



**Universidade Federal do Rio de Janeiro**

**Instituto de Química – Departamento de Química Analítica**

**Nayanna Souza Passos**

**Aplicação da Análise Multivariada para classificação de misturas de cocaína, lidocaína e  
caféina por Espectroscopia no Infravermelho com Transformada de Fourier por  
Refletância Total Atenuada (ATR-FTIR)**

**Rio de Janeiro**

**2019**

Nayanna Souza Passos

**APLICAÇÃO DA ANÁLISE MULTIVARIADA PARA CLASSIFICAÇÃO DE  
MISTURAS DE COCAÍNA, LIDOCAÍNA E CAFEÍNA POR ESPECTROSCOPIA NO  
INFRAVERMELHO COM TRANSFORMADA DE FOURIER POR REFLETÂNCIA  
TOTAL ATENUADA (ATR-FTIR)**

Trabalho de Conclusão de Curso apresentado  
ao Instituto de Química da Universidade  
Federal do Rio de Janeiro, como requisito  
parcial para obtenção do grau de Química com  
Atribuições Tecnológicas.

Orientadores:

Profa. Dra. Paula Fernandes de Aguiar

Prof. Dr. Vinicius Tadeu Kartnaller Montalvão

PCF MSc. Ana Luiza Barbosa de Oliveira

**Rio de Janeiro**

**2019**

### CIP – Catalogação na Publicação

P289a	<p>Passos, Nayanna Souza</p> <p>APLICAÇÃO DA ANÁLISE MULTIVARIADA PARA CLASSIFICAÇÃO DE MISTURAS DE COCAÍNA, LIDOCAÍNA E CAFEÍNA POR ESPECTROSCOPIA NO INFRAVERMELHO COM TRANSFORMADA DE FOURIER POR REFLETÂNCIA TOTAL ATENUADA (ATR-FTIR) / Nayanna Souza Passos. – Rio de Janeiro, 2019.</p> <p>61 pp.</p> <p>Orientadores: Paula Fernandes de Aguiar, Vinicius Tadeu Kartnaller Montalvão e Ana Luiza Barbosa de Oliveira</p> <p>Trabalho de conclusão de curso (graduação) – Universidade Federal do Rio de Janeiro, Instituto de Química, Química com Atribuições Tecnológicas, 2019.</p> <p>1. Análise por Componentes Principais. 2. Quimiometria. 3. Cocaína. 4. Adulterantes. 5. ATR-FTIR. I. Aguiar, Paula Fernandes de., orient. II. Montalvão, Vinicius Tadeu Kartnaller, orient. III. Oliveira, Ana Luiza Barbosa de, orient. IV. Título.</p>
-------	--

*Dedico este trabalho aos meus dois benzinhos: a mamãe e o Caucau, que suportaram meus momentos de estresse e ansiedade, sempre com palavras de ânimo e coragem.*

## AGRADECIMENTOS

Primeiramente, gostaria de agradecer a Deus, o Autor e consumidor da minha fé, Aquele que me fortalece, me guarda e cuida de mim. Porque Dele, por Ele e para Ele são todas as coisas. Agradeço porque, como um Pai de amor e perfeito, nunca desistiu de mim nos meus momentos mais difíceis, me fortaleceu e me preparou para passar por tudo que passei para chegar até aqui. Se não fosse Ele me conduzindo, eu nunca teria vindo parar na Química, tampouco chegaria ao término desta faculdade e, em especial, deste trabalho.

Agradeço aos meus pais, Fátima e Fernando, por terem sido tão compreensivos por eu cursar um período a mais, me apoiarem em toda esta longa e árdua caminhada e me acalmarem nos meus momentos de desespero. Ao meu pai, por todo o auxílio financeiro e incentivo em minhas escolhas profissionais, por me ensinar a disciplinar o meu tempo e não perder o foco quando os resultados não são satisfatórios, porque tudo serve para nosso “aperfeiçoamento moral e intelectual”, mas, principalmente, à minha mãe, meu braço direito, a “coronela”, o principal motivo e estímulo para que eu cresça, seja bem-sucedida, corra atrás dos meus sonhos, alcance patamares que eu mesma nunca imaginei e, como ela sempre diz: “esteja entre os grandes para ajudar os pequenos”, ou seja, reverta todo o meu conhecimento e as minhas habilidades em prol dos menos favorecidos e que mais precisam.

Agradeço também aos meus *paidrinhos*, sim, *paidrinhos*, Marly e Antônio, pois são como uma segunda mãe e um segundo pai para mim. Obrigada pelas orações, pelos conselhos, pelo carinho, amor e dedicação – sempre reservando tempo para mim, mesmo sendo tão solicitados – como padrinhos escolhidos a dedo por mim e, principalmente, por Deus para mim.

Agradeço ao meu benzinho, Cauê, por estar ao meu lado na minha rotina complicada, sempre com amor e persistência. É maravilhoso ver o que Deus tem feito em nossas vidas e como Ele nos sustentou para chegarmos até aqui. Que Ele nos conduza para, assim como Abraão, sairmos de Harã e irmos em direção à Canaã que Ele tem para nós.

Agradeço a todos que, de alguma forma, contribuíram para este trabalho. Minha muito amada orientadora professora Paula Fernandes – a que não canta – por aceitar me orientar mesmo antes de entender direito o que eu queria fazer; por estar sempre disposta a tirar minhas dúvidas e me dar todos os documentos que sempre pedia, mesmo em fase de recuperação do joelho; sempre muito paciente para comigo. Ao meu orientador Vinícius Kartnaller, sempre um anjo de bondade, mediador, disposto a me explicar com riqueza de detalhes tudo o que precisava saber sobre as partes teóricas mais complexas do trabalho, por sempre abrir a sua agenda para me atender e abraçar este projeto, mesmo já em andamento. Não viu dia, nem hora, nem final

de semana, tampouco feriado e sempre respondia as minhas muitas perguntas com paciência e ânimo.

Sou extremamente grata aos meus chefes da Polícia Federal (NUCRIM/SETEC/SR/DPF/RJ), que fizeram com que meu estágio fosse muito mais produtivo, tranquilo e divertido: Peritos Criminais Federais (PCFs) Ana Luíza, Carol Possato, Alexandre Gonzaga, Cleverson, Sérgio, Perruso, Emílio, Torres, Peçanha e Figueiredo. À minha orientadora Ana Luiza, uma atleta completa, de cuja dissertação o meu estudo foi originado, pela concessão de muitos materiais de estudo e conversas enriquecedoras para a melhoria deste trabalho. À minha supervisora Carol Possato, que foi uma mãe para mim – no sentido amplo, de puxar a orelha também – e sempre se preocupou comigo, conversava com carinho e, assim como os demais peritos do laboratório, não mantinha uma relação vertical por sua posição hierárquica superior. Ao meu irmão do coração, Alexandre Gonzaga, o químico mais raiz do laboratório, por ser o sonho de profissional em Química que eu quero ser “quando crescer”. Ao meu “pai do lab”, Cleverson, pelas conversas no café da manhã, por tanto aprendizado e cuidado por mim, sempre se preocupando e torcendo pelo meu futuro, com bom humor e muito incentivo. Ao Sérgio por sempre me instigar a pensar com tantas perguntas-surpresa e pelos maravilhosos livros de Química Forense e Técnicas Analíticas, que somaram muito a este trabalho e me trouxeram um grande conhecimento teórico do que desenvolvi ao longo do estágio. Ao Perruso, que mesmo com uma passagem pelo Rio de Janeiro tão breve, me ensinou bastante e trouxe muita luz ao laboratório. E ao professor Carlos Riehl por me supervisionar neste estágio e estar sempre solícito para assinar as papeladas, mesmo imediatamente após a virada do ano – além das brilhantes aulas de Criminalística e Toxicologia, que só fizeram com que eu amasse ainda mais a área da Química Forense. Aos amigos Emílio, Torres, Emerson, Figueiredo, pelas boas conversas e hospitalidade.

E por fim, mas não menos importante, agradeço a todos os familiares e amigos que foram essenciais para a minha caminhada na faculdade e na vida. Aos meus irmãos, principalmente Ricardo e Raniere, pelas orações e palavras de ânimo a todo momento. Ao meu amigo Wilken, por ser meu companheiro de muitas matérias, estudar comigo e me incentivar nas minhas loucuras de mil disciplinas por período; ao meu amigo Marcos Mesquita, que por muitos períodos só me mandava ir dormir, mas sempre foi meu irmãozinho rabugento preferido, cuidando de mim; a Vanessa Farelo, pelos cadernos impecáveis e pela amizade – não me esqueci de ressaltar os cadernos! -; a todos os membros do LaQuiS, super iluminados e cheios de amor, que viraram uma família para mim; aos meus amigos Luiara, José, Vinício Francisco, Marcella Tourinho, Fernanda Mardegan, Anderson Granata e Marcelo. Aos professores que

deixaram sua marca na minha formação: Sérgio Machado, Fernanda Arruda, Carlos Riehl, João Cajaíba, Carlos Eduardo, Cláudio Cerqueira e Rosângela Cappella, Elis Eleutherio, Henrique Marcelo, Fábio Azamor (e toda a maravilhosa equipe da Triagem IX). E se esqueci alguém, agradeço a tantos outros que passaram pela minha trajetória, levaram um pouco de mim e deixaram um pouco de si.

Nayanna Souza Passos

*“Se projetas alguma coisa, ela te sairá bem; e a luz brilhará em seus caminhos”. (Jó 22:28)*



**RESUMO**  
**PROJETO DE CURSO**

**TÍTULO:** APLICAÇÃO DA ANÁLISE MULTIVARIADA PARA CLASSIFICAÇÃO DE MISTURAS DE COCAÍNA, LIDOCAÍNA E CAFEÍNA POR ESPECTROSCOPIA NO INFRAVERMELHO COM TRANSFORMADA DE FOURIER POR REFLETÂNCIA TOTAL ATENUADA (ATR-FTIR)

**ALUNA:** Nayanna Souza Passos

**ORIENTADORES:** Profa. Paula Fernandes de Aguiar – DQA – Instituto de Química – UFRJ e Prof. Vinicius Tadeu Kartnaller Montalvão – DQA – Instituto de Química – UFRJ

**COLABORADORA:** Msc. Ana Luiza Barbosa de Oliveira – Polícia Federal do Estado do Rio de Janeiro.

**DATA DA DEFESA:** 07/08/2019

**PALAVRAS-CHAVE:** Análise por Componentes Principais, Quimiometria, Cocaína, Adulterantes, ATR-FTIR.

O tráfico de drogas é uma grave ameaça à segurança e saúde pública. Para combatê-lo, são necessárias ferramentas que auxiliem a Inteligência Policial, como a Química Forense. A cocaína é um exemplo de substância proscrita a respeito da qual estão sendo feitos vários estudos que objetivam auxiliar a Polícia na luta contra o narcotráfico. Em alguns desses trabalhos recentes, foi verificada a presença de lidocaína e cafeína como dois dos principais agentes de adulteração da cocaína comercializadas nas ruas. Portanto, o presente trabalho fará uso de uma das técnicas analíticas mais difundidas nos laboratórios de Química Forense da atualidade, a Espectroscopia no Infravermelho com Transformada de Fourier por Refletância Total Atenuada (ATR-FTIR), a fim de se estudar mais detalhadamente os sistemas binários e ternários de cocaína, lidocaína e cafeína, com o objetivo de construir um modelo que permita a classificação destas misturas, através da Análise por Componentes Principais dos espectros de infravermelho. O planejamento de preparo das amostras buscou variar a concentração de cada um dos três componentes em cerca de 10%, de modo que abrangesse o intervalo de 0-100%. As amostras foram pesadas em duplicata e cada uma das duplicatas foi analisada três vezes, calculando-se o espectro médio de cada amostra. Estes espectros médios sofreram processos de pré-processamento geral (centralização pela média e escalamento por variância) e foram submetidos à Análise por Componentes Principais através do programa de estatística. O número ótimo de Componentes Principais foi de 3 e a variância relativa total explicada do sistema foi de 90,98%, porém o modelo não correspondeu à estrutura original dos dados. Por fim, foi feita a transformação por primeira derivada da matriz original de dados, após o pré-processamento geral, resultando em uma representação gráfica condizente à composição das amostras que puderam, então, ser classificadas.

Autorizamos a divulgação no Repositório Institucional da UFRJ-Pantheon.

Aluno: \_\_\_\_\_

Orientadora: \_\_\_\_\_

Assinatura

Assinatura

## LISTA DE FIGURAS

Figura 1 - Espectro eletromagnético, separado nas regiões de acordo com o comprimento de onda e a frequência correspondentes.....	6
Figura 2 - Representação gráfica do princípio da Refletância Total Atenuada.....	7
Figura 3 –Três variáveis de um sistema qualquer foram medidas N vezes e entre elas foi vista uma relação (dependência linear).....	12
Figura 4 – Variância residual da matriz <b>X</b> à medida que se processam novos componentes principais. ....	13
Figura 5 – Representação geométrica do significado de <i>scores</i> .....	14
Figura 6 – Gráfico de <i>scores</i> da Análise por Componentes Principais (PCA) de 91 amostras de cocaína nas suas formas de sal (cloridrato) e base, com a presença de dois adulterantes: cloridrato de lidocaína, cafeína e dois diluentes: ácido benzoico e ácido bórico, analisada por ATR-FTIR.....	15
Figura 7 – Gráfico 1-D de <i>loadings</i> da Análise por Componentes Principais (PCA) de amostras de cocaína de rua do estado de Minas Gerais, na presença dos adulterantes: cafeína, lidocaína e fenacetina, analisadas por Cromatografia Gasosa acoplada a Espectrometria de Massas (CG-EM). ....	16
Figura 8 – Gráfico 2-D de <i>loadings</i> da Análise por Componentes Principais do mesmo sistema, também analisado por CG-EM, de 52 amostras de cocaína “de rua” do Rio de Janeiro. Este gráfico mostra a contribuição de cada variável para PC1 e PC2.....	17
Figura 9 – Fluxograma explicativo da estratégia experimental adotada no presente trabalho. O início e o fim estão destacados, para melhor esclarecimento.....	19
Figura 10 – Balança microanalítica Mettler Toledo XPE, com resolução de 10 µg, utilizada para pesagem das amostras preparadas. ....	20
Figura 11 – Espectrofotômetro Nicolet 380 FT-IR equipado com acessório de refletância total atenuada (ATR) Smart Orbit, usando um cristal de diamante como elemento de reflexão interna. ....	21
Figura 12 - Locais do eppendorf a partir dos quais se tiraram alíquotas de cada duplicata para que fossem analisadas por ATR-FTIR. ....	21
Figura 13 - Espectros de todas as seis réplicas de uma das amostras e seu espectro médio correspondente. ....	23
Figura 14 – Espectros de infravermelho de cocaína, na forma de cloridrato, cloridrato de lidocaína e cafeína puras, na região de 4000-500 cm <sup>-1</sup> , coletados pelo espectrofotômetro Nicolet 380 ATR-FTIR.	

As regiões em destaque foram as primeiras a serem excluídas, devido à alta variação aleatória presente (ruído).....	24
Figura 15 – Espectro de infravermelho na região de trabalho (1850-500 $\text{cm}^{-1}$ ) do cloridrato de cocaína. As letras correspondem às bandas assinaladas na Tabela 3. ....	25
Figura 16 - Estrutura química da cocaína na forma de cloridrato .....	25
Figura 17 - Espectro de infravermelho na região de trabalho (1850-500 $\text{cm}^{-1}$ ) do cloridrato de lidocaína. As letras correspondem às bandas assinaladas na Tabela 4. ....	26
Figura 18 - Estrutura química da lidocaína na forma de cloridrato.....	26
Figura 19 - Espectro de infravermelho na região de trabalho (1850-500 $\text{cm}^{-1}$ ) da cafeína. As letras correspondem às bandas assinaladas na Tabela 5. ....	27
Figura 20 - Estrutura química da cafeína .....	27
Figura 21 – Espectros de todas as amostras analisadas, coletados pelo espectrofotômetro Nicolet 380 FT-IR, na região de 1850-500 $\text{cm}^{-1}$ .....	28
Figura 22 – Variâncias residual e explicada da matriz de dados à medida que se processa a Análise por Componentes Principais, após a aplicação de pré-processamentos gerais.....	29
Figura 23 - Variância residual absoluta da matriz de dados e a evolução do seu decréscimo conforme são calculadas as Componentes Principais.....	30
Figura 24 - Gráfico de <i>scores</i> PC1 vs. PC2, que destaca o grupo de misturas binárias, e seus padrões correspondentes, de lidocaína e cafeína ao longo da PC2.....	31
Figura 25 - Gráfico de <i>scores</i> PC1 vs. PC3, que destaca o grupo de misturas binárias, e seus padrões correspondentes, de cocaína e cafeína ao longo da PC3. ....	32
Figura 26 – Gráfico de <i>scores</i> PC3 vs. PC2, indicando uma relação entre as amostras similar à de um diagrama de três componentes, tornando a possibilidade de classificação por composição promissora. ....	33
Figura 27 – Gráfico de <i>loadings</i> PC1 vs. PC2.....	34
Figura 28 – Gráfico de <i>loadings</i> para a primeira Componente Principal (PC1) .....	34

Figura 29 - Gráfico de <i>loadings</i> para a segunda Componente Principal (PC2) .....	35
Figura 30 - Gráfico de <i>loadings</i> para a terceira Componente Principal (PC3) .....	35
Figura 31 - Variâncias residual e explicada da matriz de dados à medida que se processa a Análise por Componentes Principais, após a aplicação de pré-processamentos gerais e da transformação por primeira derivada. ....	36
Figura 32 - Gráfico de <i>scores</i> PC1 vs. PC2, para os dados pré-processados por: auto escalamento e transformada por primeira derivada. É nítida a possível classificação por composição, de acordo com o diagrama de três componentes. ....	37
Figura 33 – Gráfico de <i>loadings</i> da PC1 dos dados pré-processados por auto escalamento e transformação por primeira derivada, refletindo a uniformidade de importância das variáveis no modelo de PCA construído. ....	41
Figura 34 - Gráfico de <i>loadings</i> PC1 vs. PC2 para os dados pré-processados por primeira derivada ..	41

## LISTA DE TABELAS

Tabela 1 – Técnicas analíticas utilizadas nos laboratórios de Química Forense, classificados segundo o seu poder discriminatório. ....	4
Tabela 2 – Técnicas de pré-processamento mais utilizadas. ....	18
Tabela 3 - Interpretação dos fenômenos químicos vibracionais que resultam nas bandas de maior intensidade (indicadas na Figura 15) do espectro da cocaína na forma de cloridrato na região de 1850-500 $\text{cm}^{-1}$ . ....	25
Tabela 4 - Interpretação dos fenômenos químicos vibracionais que resultam nas bandas de maior intensidade (indicadas na Figura 17) do espectro da lidocaína na forma de cloridrato na região de 1850-500 $\text{cm}^{-1}$ . ....	26
Tabela 5 - Interpretação dos fenômenos químicos vibracionais que resultam nas bandas de maior intensidade (indicadas na Figura 19) do espectro da lidocaína na forma de cloridrato na região de 1850-500 $\text{cm}^{-1}$ . ....	27
Tabela 6 – Variância explicada por cada PC calculado e a cumulativa correspondente para o sistema, após aplicação dos métodos de pré-processamentos gerais. Em destaque, o número ótimo de PCs definido para regra empírica. ....	30
Tabela 7 - Variância explicada por cada PC calculado e a cumulativa correspondente para o sistema, após aplicação dos métodos de pré-processamentos gerais e por primeira derivada. Em destaque, o número ótimo de PCs definido para regra empírica. ....	37
Tabela 8 – Resultado da classificação das amostras, a partir do gráfico de <i>scores</i> PC2 vs. PC1 dos dados pré-processados por auto escalamento e transformação por primeira derivada. As amostras que apresentaram uma classificação errada estão em destaque. ....	38

## LISTA DE SIGLAS

ALS	<i>Alternating Least Squares</i> (Mínimos Quadrados Alternativos)
ATR	<i>Attenuated Total Reflectance</i> (Refletância Total Atenuada)
ATR-FTIR	<i>Attenuated Total Reflectance – Fourier Transform Infrared Spectroscopy</i> (Espectroscopia no Infravermelho com Transformada de Fourier por Refletância Total Atenuada)
CG-EM	Cromatografia Gasosa acoplada a Espectrometria de Massas
DPF	Departamento de Polícia Federal
FTIR	<i>Fourier Transform Infrared Spectroscopy</i> (Espectroscopia no Infravermelho com Transformada de Fourier)
IV	Infravermelho
MSC	<i>Multiplicative Scatter Correction</i> (Correção do Efeito Multiplicativo por Espalhamento)
NUCRIM	Núcleo de Criminalística
PC	<i>Principal Component</i> (Componente Principal)
PCA	<i>Principal Component Analysis</i> (Análise por Componentes Principais)
PCF	Perito(a) Criminal Federal
PCR	<i>Principal Component Regression</i> (Regressão por Componentes Principais)
PLS	<i>Partial Least Squares</i> (Mínimos Quadrados Parciais)
RSD	<i>Relative Standard Deviation</i> (Desvio Padrão Relativo)
SETEC	Setor Técnico Científico
SNV	<i>Standard Normal Variate</i> (Variável Normal Padrão)
SR	Setor Regional
TRV	<i>Total Residual Variance</i> (Variância Residual Total)
UNODC	<i>United Nations Office of Drugs and Crime</i> (Escritório das Nações Unidas sobre Drogas e Crime)
UV	Ultravioleta

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
<b>2</b>	<b>OBJETIVO</b>	<b>3</b>
<b>3</b>	<b>FUNDAMENTOS TEÓRICOS</b>	<b>4</b>
3.1	Métodos de análise de cocaína	4
3.2	Espectroscopia no Infravermelho	5
3.2.1	<i>Espectroscopia no Infravermelho com Transformada de Fourier por Refletância Total Atenuada (ATR-FTIR)</i>	6
3.2.2	<i>Lei de Lambert-Beer</i>	7
3.3	Análise Multivariada	8
3.4	Quimiometria	9
3.5	Análise por Componentes Principais (PCA)	10
3.5.1	<i>Conceitos Fundamentais</i>	10
3.5.2	<i>Entendimento geométrico do PCA</i>	11
3.5.3	<i>Número de Componentes Principais</i>	13
3.5.4	<i>Scores e loadings</i>	14
3.5.4.1	<i>Interpretação gráfica dos scores</i>	15
3.5.4.2	<i>Interpretação gráfica de loadings</i>	15
3.6	Pré-Processamento de dados	17
<b>4</b>	<b>ESTRATÉGIA EXPERIMENTAL</b>	<b>19</b>
<b>5</b>	<b>METODOLOGIA</b>	<b>20</b>
5.1	Avaliação da repetibilidade do método e estabilidade do equipamento	20
5.1.1	<i>Preparo das amostras</i>	20
5.1.2	<i>Análise por ATR-FTIR</i>	20
5.2	Avaliação da região do espectro a se trabalhar	21
5.3	Aplicação dos pré-processamentos gerais para os espectros	21
5.4	Análise por Componentes Principais preliminar	22
5.5	Novo pré-processamento e Análise por Componentes Principais	22
<b>6</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>23</b>
6.1	Avaliação da repetibilidade do método e da estabilidade do equipamento	23
6.2	Avaliação da região do espectro a se trabalhar	24
6.2.1	<i>Interpretação química dos espectros dos compostos puros</i>	25
6.3	Aplicação dos pré-processamentos gerais para os espectros	28
6.4	Análise por Componentes Principais preliminar	29
6.4.1	<i>Análise da variância residual</i>	29
6.4.2	<i>Análise dos scores</i>	31

6.4.3	<i>Análise dos loadings</i> .....	33
<b>6.5</b>	<b>Novo pré-processamento e Análise por Componentes Principais</b> .....	35
6.5.1	<i>Características da técnica de pré-processamento escolhida</i> .....	35
6.5.2	<i>Análise da variância residual</i> .....	36
6.5.3	<i>Análise de scores</i> .....	37
6.5.4	<i>Análise de loadings</i> .....	41
<b>7</b>	<b>CONCLUSÃO</b> .....	42
<b>8</b>	<b>PERSPECTIVAS FUTURAS</b> .....	43
	<b>REFERÊNCIAS</b> .....	44



## 1 INTRODUÇÃO

O tráfico de drogas é uma grave ameaça à segurança e à saúde públicas. À segurança, por financiar arsenais de grupos criminosos, promovendo a violência (OLIVEIRA, 2011); à saúde, por disseminar substâncias psicotrópicas que podem causar dependência física e psíquica, gerando um crescente gasto público com tratamento médico e internação hospitalar (SILVA, 2016).

O combate a esta prática ilícita, de abrangência mundial, é um trabalho árduo, que demanda diversas ferramentas de inteligência policial, muita pesquisa e investimento, pois as diversas leis criadas, de cunho proibitivo e punitivo, não foram capazes de contê-lo. Portanto, fez-se necessário que houvesse uma nova abordagem ao problema.

Neste quesito, a Química Forense, com a produção de provas materiais objetivas, é uma ferramenta valiosa para que se analisem amostras suspeitas das mais variadas origens (bebidas, drogas ilícitas, explosivos, combustíveis, agrotóxicos) de modo que os resultados sejam confiáveis, imparciais e consigam, no caso do tráfico de entorpecentes, auxiliar na identificação de rotas de tráfico e origem da droga.

A cocaína, por exemplo, já foi a segunda droga mais consumida no mundo (SILVA, 2016); porém, conforme o Relatório Mundial de Drogas, ela já passou a ocupar a quinta posição de consumo em 2017 e foi a segunda substância proscrita mais apreendida no mesmo ano (UNODC, 2019). Isto representa um avanço, e um estímulo para que novos estudos sejam somados no combate estratégico à esta e outras drogas.

Através da análise química de amostras apreendidas, foi possível perceber que, no intuito de aumentar o volume da droga e, principalmente, os lucros, a cocaína passa por processos de adulteração (adição de outros compostos farmacologicamente ativos) e diluição (adição de substâncias farmacologicamente inertes) antes da venda ao usuário final, uma prática que compromete ainda mais a saúde de seus usuários, podendo levar ao óbito (BERNARDO, 2003; LAPACHINSKE, 2015; RODRIGUES, 2013).

Embora o uso de determinados adulterantes mais frequentes não siga nenhum padrão definido, podendo variar ao longo do tempo e da região, dependendo da disponibilidade e dos preços de mercado das áreas correspondentes, assim como da forma química da cocaína à qual se é adicionada (cloridrato ou cocaína base) (SANT'ANA, 2019), diversos estudos revelaram algumas tendências para amostras apreendidas em diversos estados do Brasil (BERNARDO, 2003; MALDANER, 2016; MARCELO, 2015; SANT'ANA, 2019), apontando para a lidocaína e a cafeína como principais adulterantes utilizados. Souza *et al.* (2016) classificaram amostras de cocaína de acordo com o ano e a região de apreensão a partir de seus perfis químicos,

enquanto que Sant'Ana *et al.* (2019) conseguiu correlacionar alguns perfis químicos de cocaína a determinados grupos criminosos.

Portanto, o presente trabalho fez uso de uma das técnicas analíticas mais difundidas nos laboratórios de Química Forense da atualidade, a Espectroscopia no Infravermelho com Transformada de Fourier por Refletância Total Atenuada (ATR-FTIR), a fim de se estudar mais detalhadamente os sistemas ternários de cocaína, na forma de cloridrato, e seus adulterantes: lidocaína e cafeína.

## **2 OBJETIVO**

Construir um modelo que permita a classificação do sistema de misturas binárias e ternárias de cocaína (cloridrato) e dois de seus principais adulterantes: cloridrato de lidocaína e cafeína, através da Análise por Componentes Principais (PCA) de espectros de infravermelho com Transformada de Fourier por Refletância Total Atenuada (ATR-FTIR).

### 3 FUNDAMENTOS TEÓRICOS

#### 3.1 Métodos de análise de cocaína

A análise rotineira de drogas costuma seguir os métodos recomendados para a identificação e análise de cocaína do Escritório das Nações Unidas sobre Drogas e Crime (UNODC) (UNODC, 2012). As diferentes técnicas analíticas utilizadas são classificadas em três categorias: A, B e C, de acordo com o seu potencial discriminatório (Tabela 1) (SWGDRUG, 2014).

Tabela 1 – Técnicas analíticas utilizadas nos laboratórios de Química Forense, classificados segundo o seu poder discriminatório (Fonte: Adaptado de SWGDRUG, 2014).

<b>Categoria A</b>	<b>Categoria B</b>	<b>Categoria C</b>
Espectroscopia no Infravermelho	Cromatografia em camada fina	Espectroscopia de fluorescência
Espectrometria de Massas	Cromatografia gasosa	Testes colorimétricos
Espectroscopia Raman	Cromatografia líquida	Ponto de fusão
Espectroscopia de Ressonância Magnética Nuclear (RMN)	Espectrometria de mobilidade iônica	Espectroscopia no ultravioleta
Difratometria de raios-X	Eletroforese capilar	Imunoensaio
	Microcristalização	

Quando uma técnica A, de maior poder discriminatório, é utilizada, recomenda-se que mais uma técnica, de qualquer categoria, seja utilizada. Caso, porém, não se possa acessar uma técnica da categoria A, o recomendável é que três diferentes técnicas sejam utilizadas.

Dentre as técnicas citadas, as mais disponíveis e amplamente difundidas nos laboratórios de Química Forense da Polícia Federal são a espectroscopia no infravermelho e a espectrometria de massas (acoplada a cromatografia gasosa, CG-EM). Destas, a CG-EM se destaca em sua eficácia por ser uma técnica tanto de identificação como de separação. Entretanto, ela apresenta inconvenientes como um procedimento de preparo de amostra que pode ser complexo, assim como alto custo e maior tempo de análise (GROBÉRIO, 2014), e do fato de ser uma técnica destrutiva.

Em contrapartida, as análises por espectroscopia vibracional (espectroscopia no infravermelho com transformada de Fourier – FTIR e Raman) são rápidas, não destrutivas, não demandam preparo de amostra – dependendo da técnica utilizada –, tampouco uso de solventes

e permitem o uso de aparelhos portáteis para análises em campo (ELIAERTS, 2017; HESPANHOL, 2019; MATERAZZI, 2017; PENIDO, 2017). Sua grande desvantagem, porém, é a capacidade limitada de identificação de compostos quando em presença de misturas. Este problema pode ser contornado com o uso de métodos quimiométricos de análise multivariada de dados, possibilitando o uso de um método analítico ecológico, seletivo e suficientemente sensível para a caracterização de drogas (HESPANHOL, 2019).

### 3.2 Espectroscopia no Infravermelho

A espectroscopia estuda a interação da luz (radiação eletromagnética) com a matéria. Ao conjunto de frequências absorvidas por determinada molécula, dá-se o nome de espectro de absorção; analogamente, o espectro de emissão refere-se às frequências emitidas (LEVINE, 2009). Esses processos de interação da radiação com a matéria são tratados de forma mais criteriosa pela mecânica quântica, o que foge do escopo deste trabalho.

De forma simples, entende-se que todas as moléculas apresentam uma energia interna total, que é um somatório das energias eletrônica, vibracional, rotacional e translacional (COATES, 2006). Quando elas interagem com a luz, esta energia pode sofrer uma transição para um nível mais alto, de valor  $h\nu$ , onde  $h$  é a constante de Planck e  $\nu$  é a frequência da radiação absorvida.

Dependendo deste valor, as transições podem ocorrer em níveis rotacionais, vibracionais ou eletrônicos (LEVINE, 2009). As transições rotacionais exigem radiação com energias menores, que correspondem a comprimentos de onda maiores (região do micro-ondas do espectro eletromagnético), enquanto que as transições vibracionais ocorrem quando a radiação incidida está na região do infravermelho, e, por fim, as transições eletrônicas ocorrem por absorção nas regiões do visível, UV e raios-X (Figura 1).

A região do infravermelho é dividida em três subregiões: infravermelho próximo ( $12800\text{-}4000\text{ cm}^{-1}$ )<sup>1</sup>, médio ( $4000\text{-}400\text{ cm}^{-1}$ ) e distante ( $400\text{-}10\text{ cm}^{-1}$ ) (CHALMERS, 2012). A região mais estudada é a do infravermelho médio (LEVINE, 2009).

Um modo normal de vibração molecular só é considerado ativo no infravermelho, dando origem a uma banda de absorção, quando ele gera uma variação no momento dipolo da molécula (CHALMERS, 2012; COATES, 2006; LEVINE, 2009).

---

<sup>1</sup> O comprimento de onda ( $\lambda$ ) e o número de onda ( $\tilde{\nu}$ ), que é a unidade utilizada nos espectros de absorção ou transmissão, dada em  $\text{cm}^{-1}$ , se relacionam de forma inversamente proporcional, isto é:  $\tilde{\nu} = \frac{1}{\lambda}$

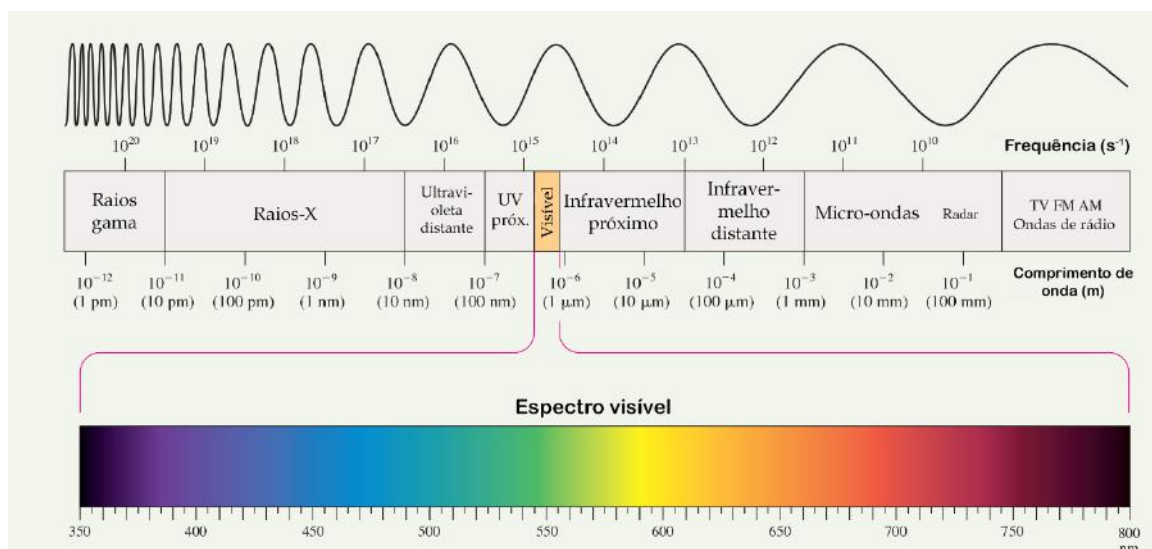


Figura 1 - Espectro eletromagnético, separado nas regiões de acordo com o comprimento de onda e a frequência correspondentes. A região da luz visível ocorre em comprimentos de onda entre 350-800 nm, conforme destacado (Fonte: Adaptado de SKOOG, 2014).

A absorção da energia da radiação por determinados grupos funcionais presentes nos compostos químicos ocorre, em geral, em regiões específicas do espectro, sendo levemente deslocadas de acordo com o seu ambiente químico – presença de outros grupos funcionais, gerando interações intra e intermoleculares (COATES, 2006; CHALMERS, 2012). Porém, ao longo do espectro no infravermelho, a região entre 1500-400 cm<sup>-1</sup> merece especial atenção por ser a região da **impressão digital**, ou seja, a que mais se diferencia de molécula para molécula. Esta região é tão sensível que pode refletir polimorfismos moleculares, grau de cristalização, estado de hidratação, entre outras características (CHALMERS, 2012).

### 3.2.1 Espectroscopia no Infravermelho com Transformada de Fourier por Refletância Total Atenuada (ATR-FTIR)

A espectrofotometria com Transformada de Fourier representou um avanço na espectroscopia no infravermelho, permitindo a detecção de sinais mais baixos e, por conseguinte, aumentando a sensibilidade do método. Um espectrofotômetro FT-IR usa um interferômetro de Michelson para formar um interferograma, que usa a transformada de Fourier para gerar um espectro de absorção no infravermelho, ao invés de dispersar radiação, como era feito pelos espectrofotômetros convencionais (LEVINE, 2009).

Um progresso ainda maior foi obtido quando os espectrofotômetros foram equipados com um acessório de Refletância Total Atenuada (ATR), que não exige preparo de amostra e mantém uma boa repetibilidade espectral (PERKINELMER, 2005).

A Refletância Total Atenuada (Figura 2) se baseia num fenômeno ótico chamado reflexão total interna, que só pode ocorrer em meios com alto índice de refração (YOUNG, 2012). Isso faz com que o feixe incidente seja refletido sucessivamente, aumentando o seu contato com a superfície da amostra (MONTALVÃO, 2013). Em regiões do espectro de infravermelho onde a amostra absorva energia, a radiação incidente será atenuada ou alterada. Essa energia alterada de cada onda incidente é passada de volta para o feixe do infravermelho, deixando o lado oposto do cristal e passando para o detector do espectrofotômetro (PERKINELMER, 2005).

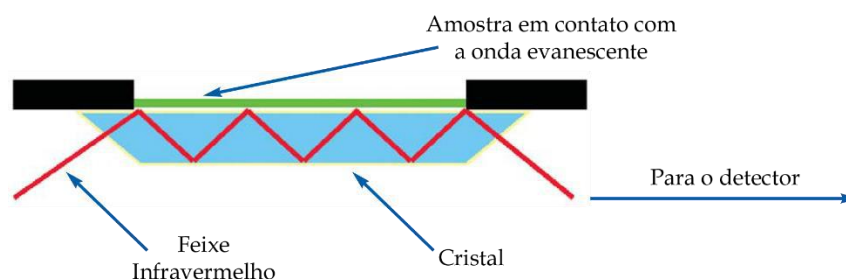


Figura 2 - Representação gráfica do princípio da Refletância Total Atenuada (Fonte: Adaptado de PERKINELMER, 2005).

O cristal mais usado é o de diamante, devido à sua robustez e durabilidade, embora o seu custo inicial seja mais caro que as demais opções (PERKINELMER, 2005). Outro inconveniente associado ao seu uso é que ele absorve na região de  $2100-1900\text{ cm}^{-1}$ , impossibilitando a extração de informações acerca da amostra analisada nessa região.

### 3.2.2 Lei de Lambert-Beer<sup>2</sup>

A espectroscopia no infravermelho, embora apresente uma aplicação qualitativa direta, também pode ser usada para análises quantitativas, através da relação da absorvância de uma amostra com a sua concentração, que é dada pela lei de Lambert-Beer (Equação 1).

$$A = b \epsilon c \quad (\text{Equação 1})$$

A absorvância (A) é diretamente proporcional à concentração da amostra (c) medida em determinado comprimento de onda  $\lambda$ .  $\epsilon$  é a constante de absortividade molar de determinada espécie neste mesmo  $\lambda$  e b é o caminho ótico que a luz percorre na amostra.

<sup>2</sup> As principais referências desta subseção estão em Esbensen (2018) e Levine (2009).

Os espectrofotômetros antigos costumavam coletar os dados no modo de transmitância (T), enquanto que atualmente, o mais usual é coletá-los no modo de refletância (R). Ambos se relacionam com a absorvância logaritmicamente (Equação 2).

$$A = -\log_{10} \left( \frac{1}{T} \right) = -\log_{10} \left( \frac{1}{R} \right) \quad (\text{Equação 2})$$

Tanto a transmitância quanto a refletância são dadas pela razão entre a luz transmitida e a incidida; porém a transmitância é medida percentualmente e a refletância é medida em escala de 0-1.

### 3.3 Análise Multivariada

A análise multivariada é uma ferramenta que reflete as características do mundo no qual estamos inseridos. As variáveis de um sistema qualquer não podem ser consideradas isoladamente quando se busca fazer uma análise mais completa, a não ser quando elas são independentes.

Desde algo com que se tenha mais contato, como o clima, que é influenciado por fatores como pressão atmosférica, temperatura, umidade, ventos (ESBENSEN, 2018); como questões mais abstratas, a exemplo do que leva uma pessoa a fazer uso de substâncias psicoativas – traumas de infância, ambiente familiar desestruturado, ambiente escolar hostil, disponibilidade da droga e outros (UNODC, 2018) – refletem como muitos sistemas são complexos e não podem ser analisados por uma única variável. Isto não significa que é impossível ou pouco relevante o estudo de um problema de forma univariada, mas revela a limitação deste tipo de abordagem.

Contudo, o fato de que se meçam muitas variáveis não implica, por si só, que se obterá sucesso ao se explorar um sistema qualquer, pois é necessário que as informações contidas nos dados obtidos sejam relevantes e busquem atender o objetivo para o qual a análise foi proposta. De forma sucinta, “é mais importante saber quais variáveis foram medidas do que quantas” (ESBENSEN, 2018, p. 38).

Traduzindo o que foi dito para uma linguagem quimiométrica, as observações ligadas a uma amostra – podendo ser o número de medições, amostras ou experimentos – são chamadas de **objetos** e as medidas obtidas a partir do número de objetos são as **variáveis** de um sistema (ESBENSEN, 2018; WOLD, 1987).

Seguindo a analogia da equação da reta  $y = ax + b$ , as variáveis podem ser independentes (variáveis  $x$ ) ou dependentes (variáveis  $y$ ). O conjunto de equações que descreve



o sistema forma por um espaço experimental, descrito pela matriz  $\mathbf{X}$ , e um espaço resposta, descrito pela matriz  $\mathbf{Y}$ .

Portanto, devido à relação de dependência existente entre as variáveis  $X$  e  $Y$ , é possível caracterizar indiretamente uma propriedade  $Y$  a partir de medições de  $X$ . Quando esta relação é linear, ao se mudar o valor de  $X$ , o valor de  $Y$  muda proporcionalmente, com uma dada incerteza.

### 3.4 Quimiometria

A Quimiometria é uma área interdisciplinar que abrange a análise estatística multivariada, a modelagem matemática, a ciência computacional e a química analítica, aplicável para (1) otimizar procedimentos de medida e experimentos; (2) extrair o máximo de informação química de um conjunto de dados analisados (BOVENS, 2019; GEMPERLINE, 2006). É uma importante ferramenta estatística que permite a melhor compreensão de dados químicos e reconhecimento de relações complexas do sistema que de outra forma passariam despercebidas (MURO, 2016). Isto pode ser imprescindível para redução de dados e previsões analíticas (SWARBRICK, 2012).

Os métodos quimiométricos mais utilizados podem ser divididos em três categorias: (a) métodos matemáticos de pré-tratamento, que organizam a informação atribuída à variação química; (b) métodos qualitativos, que dividem as amostras em classes de acordo com a similaridade dos dados espectroscópicos; (c) métodos quantitativos, também chamados de métodos de calibração, que associam dados espectrais com propriedades quantificáveis da amostra (MURO, 2016).

Alguns métodos matemáticos serão abordados mais à frente, na subseção de *Pré-Processamento*. Os métodos qualitativos se atêm apenas à matriz  $\mathbf{X}$ ; enquanto os métodos quantitativos exploram a relação entre as matrizes do espaço experimental e da resposta,  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente.

Há muitas técnicas de análise multivariada disponíveis para uso. Porém, a escolha de um método está intimamente ligada com o tipo de resposta que se deseja obter. Para tanto, é imprescindível que se trace um objetivo suficientemente claro e que os dados estejam organizados de modo que este objetivo seja alcançado (ESBENSEN, 2018).

### 3.5 Análise por Componentes Principais (PCA)<sup>3</sup>

A PCA é uma das principais ferramentas para a análise multivariada e, particularmente, para a quimiometria (GEMPERLINE, 2006). Sua história teve início em 1901, por Karl Pearson, no ramo da Estatística (PEARSON, 1901), evoluiu para sua forma mais moderna com Hotelling (HOTELLING, 1933) e alcançou a Química em torno de 1960, com Malinowski (WOLD, 1987).

Na linguagem algébrica, essa análise é responsável por decompor uma matriz  $\mathbf{X}$  na parte “estruturada” – que contém informações relevantes,  $\mathbf{TP}^T$  – e na parte ruidosa,  $\mathbf{E}$  (Equação 3) através de uma projeção ortogonal que permite a visualização de dados escondidos, ou **latentes**, em  $\mathbf{X}$ .

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (\text{Equação 3})$$

$\mathbf{X}$  é uma matriz  $n \times p$ , onde as linhas  $n$  são os objetos e as colunas  $p$  são as variáveis independentes. A parte matemática não será explorada a fundo no presente texto, a não ser quando estritamente necessário. Porém, pode ser encontrada com riqueza de detalhes em Jolliffe (2002).

A fim de que se tenha um entendimento claro do que vai ser abordado a seguir, é necessário que se recorra aos conceitos fundamentais do estudo da Estatística.

#### 3.5.1 Conceitos Fundamentais

A palavra-chave da estatística é variabilidade. As ferramentas estatísticas estudam e gerenciam a variabilidade do sistema a fim de prever esta variabilidade no futuro. Ela nem sempre representa algo ruim, dependendo do contexto e da aplicação.

Esta variabilidade, ou dispersão, representa o grau de afastamento dos valores das observações no que diz respeito à média<sup>4</sup> da distribuição (Equação 4). A dispersão pode ser medida de forma absoluta, através do desvio padrão (Equação 5) e da variância (Equação 6) ou de forma relativa, através do coeficiente de variação, também chamado de desvio padrão relativo (RSD) (Equação 7) (GEMPERLINE, 2006).

---

<sup>3</sup> Este tópico foi majoritariamente advindo de Esbensen (2018), sendo essa uma excelente fonte de estudo para aprofundamento do assunto por parte do leitor. Outras referências serão dadas ao longo do texto.

<sup>4</sup> A média representa a tendência central das medições, desde que os dados sejam representativos, a sua distribuição seja normal, os resultados sejam independentes e não haja valores aberrantes.

Média de distribuição ( $\bar{x}$ ):  $\bar{x}$  representa o valor médio;  $x_i$ , os valores das medições e  $n$ , o número total de medições.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (\text{Equação 4})$$

Desvio padrão ( $s$ ):  $n - 1$  representa o número de graus de liberdade, ou de desvios independentes, determinado pela quantidade de comparações independentes que podem ser feitas entre  $n$  unidades da amostra.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (\text{Equação 5})$$

Variância ( $s^2$ ): é o quadrado do desvio padrão.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (\text{Equação 6})$$

Desvio padrão relativo (RSD):  $s$  é o desvio padrão e  $\bar{x}$  é o valor médio.

$$RSD = \frac{s}{\bar{x}} \times 100\% \quad (\text{Equação 7})$$

A variabilidade é vista como algo ruim quando está associada a erros e não à diversidade do espaço amostral. Estes erros podem ser grosseiros, sistemáticos ou aleatórios<sup>5</sup> (GEMPERLINE, 2006). Dentre esses, os erros grosseiros se destacam por poder gerar valores aberrantes – resultados muito discrepantes em relação ao valor central das medidas.

Contudo, é importante salientar que nem todos os valores aberrantes, ou *outliers*, são oriundos de erros grosseiros. Eles podem ser apenas valores que não se ajustam às mesmas características de um dado conjunto, representando um novo fenômeno.

### 3.5.2 Entendimento geométrico do PCA

Inicialmente, considera-se que todas as  $p$  variáveis são linearmente independentes; portanto, o espaço gerado pelas colunas (espaço-coluna) apresenta dimensão  $p$ . Como, porém, isto quase nunca é verdade, a real dimensão do espaço-coluna é menor que  $p$  e a dependência linear entre algumas variáveis faz com que surja uma direção de maior variância no espaço gerado pelas variáveis (espaço-coluna). A Figura 3 ilustra o que foi dito para  $p = 3$ .

---

<sup>5</sup> Os erros grosseiros são ocasionais e, em sua maioria, podem ser evitados. Os erros sistemáticos têm causas assinaláveis, são independentes do número de medições e podem ser corrigidos nos resultados. Os erros aleatórios são provenientes de flutuações do método de análise e podem ser reduzidos com o aumento do número de medições ou com o aumento da qualidade dos dados gerados.

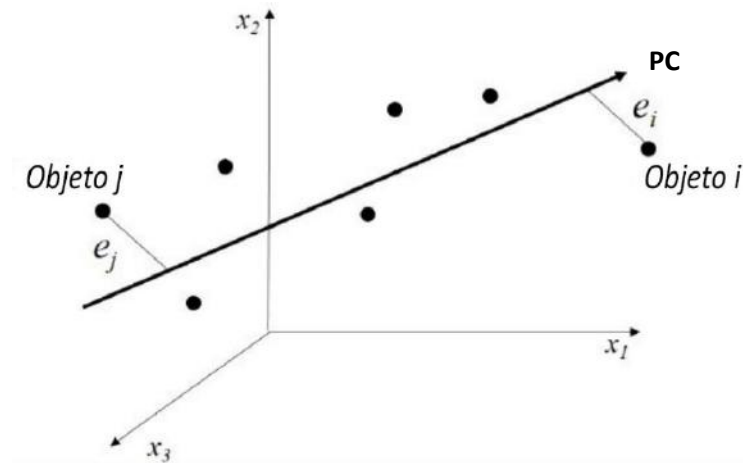


Figura 3 –Três variáveis de um sistema qualquer foram medidas  $N$  vezes e entre elas foi vista uma relação (dependência linear). Os dados apresentam uma maior dispersão ao longo do eixo que traduz esta dependência linear. Cada objeto do espaço  $p$ -dimensional será projetado ao longo deste eixo PC e a distância entre o objeto original e sua projeção é chamada de  $e_n$ , onde  $n$  são os objetos (Fonte: Adaptado de ESBENSEN, 2018, p. 157).

Esta direção é chamada de primeira Componente Principal (PC1) e é uma combinação linear de todas as variáveis  $p$ . Ao longo desse eixo, cada objeto será projetado, e a distância entre o valor original do objeto, no espaço  $p$ -dimensional, e a sua projeção ao longo do eixo PC é chamada de resíduo do objeto. Quanto maior for essa distância residual  $e_n$ , onde  $n$  são os objetos, maior será a contribuição do objeto correspondente para a direção do PC.

Conforme se aumenta a complexidade dos dados e, conseqüentemente, as relações entre as variáveis, mais eixos PC serão necessários. Por definição, o segundo componente principal (PC2) é ortogonal ao eixo PC1, a partir da mesma origem, e aponta na direção da segunda maior variância do conjunto de dados. O mesmo é verdadeiro para componentes principais de ordens superiores (PC3, PC4 e assim sucessivamente). É importante ressaltar que cada um dos componentes principais descreve uma fração da variância total do sistema menor que o anterior, ou seja,  $PC1 > PC2 > PC3 > PC4 \dots$  (Figura 4).

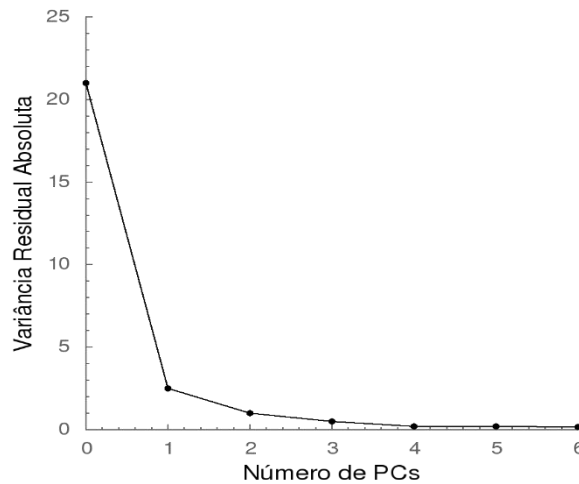


Figura 4 – Variância residual da matriz  $\mathbf{X}$  à medida que se processam novos componentes principais. O PC1 cobre a maior variância dos dados e os demais componentes principais, o que restou de variância por se explicar do componente anteriormente processado. A variância residual é uma função decrescente conforme se aumenta o número de PCs ou, alternativamente, a variância explicada é uma função crescente ao se aumentar o número de PCs (Fonte: Adaptado de ESBENSEN, 2018, p. 198).

O valor referente ao PC-0 do gráfico de variância faz menção ao valor absoluto do resíduo da matriz original, ou à variância residual total (TRV), sem a construção de um modelo de Análise por Componentes Principais. Este valor é obtido através da razão entre o somatório dos quadrados dos resíduos de cada objeto e o produto dos objetos e variáveis (Equação 8). O quadrado da variância residual de um objeto, por sua vez, é composto pelo somatório dos quadrados dos resíduos de cada variável que o compõe (Equação 9).

$$TRV = \frac{e_{total}^2}{p \times n} = \frac{\sum_{i=1}^n e_i^2}{p \times n} \quad (\text{Equação 8})$$

Quadrado da variância residual de um objeto: é a soma dos quadrados dos resíduos de cada variável  $k$  para um objeto  $i$ .

$$e_i^2 = \sum_{k=1}^p e_{ik}^2 \quad (\text{Equação 9})$$

### 3.5.3 Número de Componentes Principais

Sucintamente, pode-se dizer que a Análise por Componentes Principais é uma transformação linear que projeta os dados de uma matriz original  $\mathbf{X}$  em uma nova base ortogonal que reflete a dependência linear entre as variáveis, possuindo, assim, uma dimensão menor ou igual ao número total de variáveis do sistema. Logo, há um limite máximo para o número de componentes principais (PCs), que pode ser  $n-1$  (um valor abaixo do número total de objetos) ou  $p$  (número total de variáveis), dependendo do que for menor.

O uso do número máximo de PCs se traduz por englobar a parte da matriz original referente ao ruído (a parte E da Equação 3), o que não gera nenhum benefício prático para a

análise de dados. Há um número ótimo de componentes principais a se usar, mas é de responsabilidade do analista observar as peculiaridades do seu problema e se este número ótimo traz as respostas às perguntas que sua análise busca responder.

A regra empírica que geralmente se aplica é que quando houver uma quebra brusca de inclinação da função decrescente de variância residual – não explicada pela PCA, como ocorre no ponto do PC1 da Figura 4, este PC é um potencial número ótimo. No caso observado, a inclusão do PC2 traz um benefício mínimo ao modelo e, a partir do PC3, não ocorre mais variação. Há, também, outra regra de determinação do número ótimo de PCs, que consiste em alcançar, pelo menos, 70% de variância total explicada do sistema (GEMPERLINE, 2006).

### 3.5.4 Scores e loadings

Retornando à Equação 3, o termo  $TP^T$  pode ser desmembrado em:

$$TP^T = t_1 p_1^T + t_2 p_2^T + t_3 p_3^T + \dots + t_A p_A^T \quad (\text{Equação 10})$$

Onde cada número representa o PC correspondente e A é o número ótimo de PCs.

Os termos  $p$  do lado direito da Equação 10, chamados *loadings*, representam vetores que definem o espaço das variáveis, ou a combinação linear de cada variável e sua respectiva contribuição para cada PC.

Os termos  $t$  do lado direito da Equação 10, por sua vez, são chamados *scores* e representam o valor da coordenada de um objeto projetado no espaço gerado pelos PCs (Figura 5).

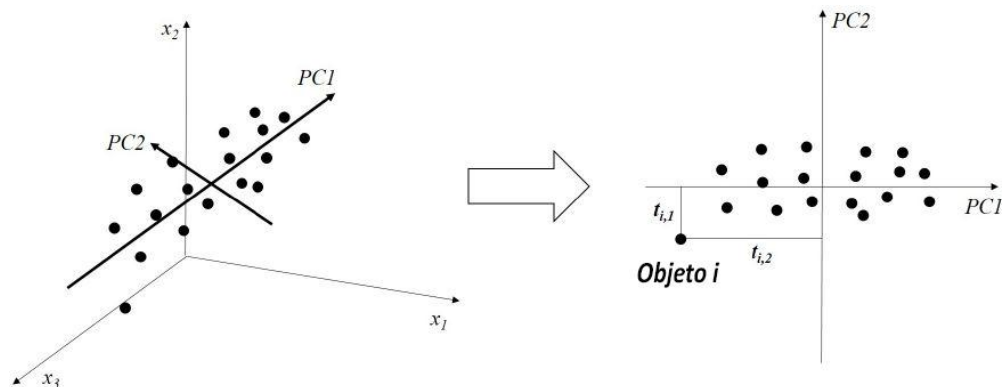


Figura 5 – Representação geométrica do significado de *scores*. Os *scores* são os valores da coordenada de cada objeto no novo espaço formado pelos Componentes Principais (PCs). Portanto, num espaço bidimensional, o objeto  $i$  apresentará uma coordenada no eixo PC1,  $t_{i,1}$ , e uma coordenada no eixo PC2,  $t_{i,2}$  (Fonte: Adaptado de ESBENSEN, 2018, p. 166).

Essa relação dos *loadings* com as variáveis e dos *scores* com os objetos pode ser visualizada graficamente.

### 3.5.4.1 Interpretação gráfica dos *scores*

Embora seja possível representar graficamente a relação entre cada PC individualmente e os respectivos objetos, o que geralmente fornece maiores informações é o estudo do comportamento entre os PCs. A fim de se exemplificar, extraiu-se o gráfico de *scores* do trabalho de Rodrigues (2013) (Figura 6).

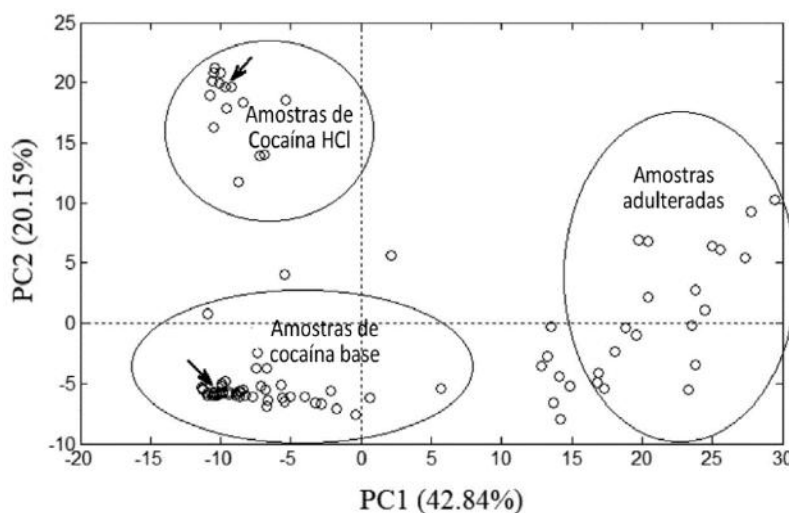


Figura 6 – Gráfico de *scores* da Análise por Componentes Principais (PCA) de 91 amostras de cocaína nas suas formas de sal (cloridrato) e base, com a presença de dois adulterantes: cloridrato de lidocaína, cafeína e dois diluentes: ácido benzoico e ácido bórico, analisada por ATR-FTIR (Fonte: Adaptado de RODRIGUES, 2013, p. 512).

É possível verificar que a Componente Principal 1 (PC1) explica 42,84% da variância do sistema, enquanto que a Componente Principal 2 (PC2), 20,15%, totalizando 62,99% da variância total do sistema. O eixo de maior variância (PC1) consegue discriminar bem as amostras em relação à sua composição, com boa separação das amostras mais adulteradas; enquanto que o segundo eixo de maior variância (PC2) faz a separação pela forma química da cocaína (cloridrato ou base).

O gráfico de *scores* auxilia na identificação de valores aberrantes, de tendências, grupos com características similares, avaliação das replicatas, entre outros aspectos. E juntamente com o gráfico de *loadings*, pode-se fazer inferências sobre as variáveis originais e sua contribuição no modelo (ESBENSEN, 2018, p. 171).

### 3.5.4.2 Interpretação gráfica de *loadings*

Os gráficos de *loadings* trazem a informação sobre a inter-relação entre as variáveis. No caso dos *loadings*, a interpretação 1-D e 2-D são igualmente comuns, sendo que na primeira representação, é possível ver quais são as variáveis que mais contribuem com cada PC,

enquanto que a segunda representação permite visualizar a relação e possíveis similaridades entre as variáveis do sistema. Os gráficos extraídos dos trabalhos de Souza (2016) (Figura 7) e de Sant'Ana (2019) (Figura 8) ilustram ambos os casos, respectivamente. Os *loadings*, em geral, são normalizados no intervalo  $[-1,1]$ .

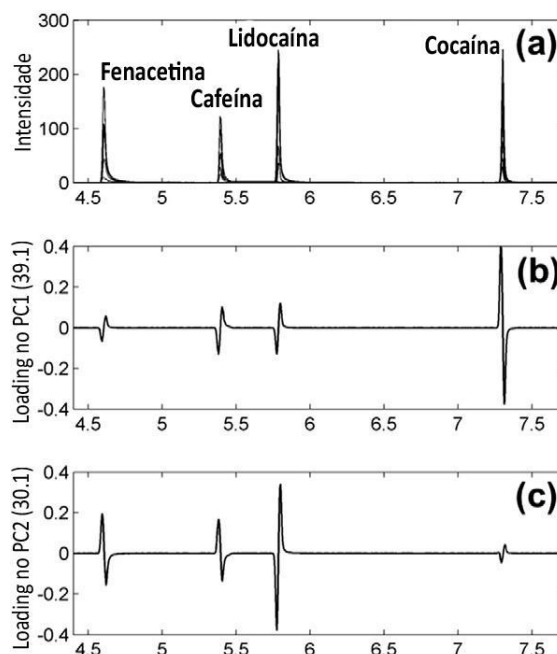


Figura 7 – Gráfico 1-D de *loadings* da Análise por Componentes Principais (PCA) de amostras de cocaína de rua do estado de Minas Gerais, na presença dos adulterantes: cafeína, lidocaína e fenacetina, analisadas por Cromatografia Gasosa acoplada a Espectrometria de Massas (CG-EM). (a) é um cromatograma representativo de uma amostra de cocaína e seus respectivos adulterantes, enquanto que (b) e (c) são os gráficos de *loadings* 1-D para o PC1 e PC2, respectivamente (Fonte: Adaptado de SOUZA, 2016, p. 77).

Através da Figura 7, é possível perceber que a PC1 explica 39,1% da variância do sistema, enquanto que a PC2, 30,1%, perfazendo uma variância total explicada de 69,2%. O gráfico 1-D apresenta o mesmo aspecto de cromatograma dos dados originais, e o mesmo se repete para outras técnicas analíticas, como a espectroscopia; por isso, é chamado de **espectro de loadings**.

É possível verificar em (b) que a cocaína é a variável que mais contribui com a primeira componente principal (PC1), enquanto que, para a PC2, a variável mais importante é a lidocaína, conforme (c).



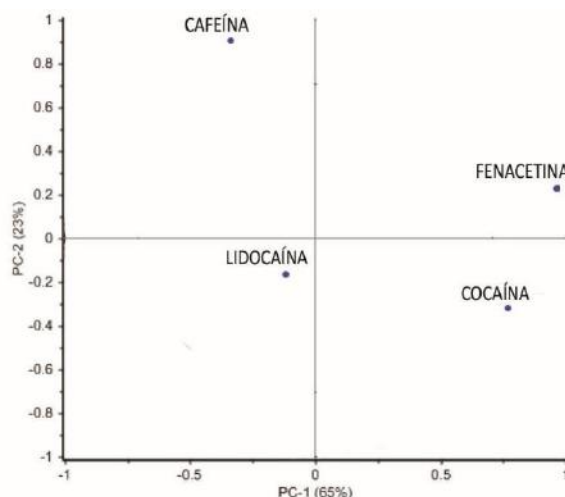


Figura 8 – Gráfico 2-D de *loadings* da Análise por Componentes Principais do mesmo sistema da Figura 7, também analisado por CG-EM, de 52 amostras de cocaína “de rua” do Rio de Janeiro. Este gráfico mostra a contribuição de cada variável para PC1 e PC2. A cocaína e a cafeína são negativamente correlacionadas pela PC1, estando em valores opostos neste eixo, enquanto que a fenacetina e a lidocaína estão negativamente correlacionadas pela PC2 (Fonte: Adaptado de SANT’ANA, 2019, p. 383).

Os gráficos 2-D de *loadings* mostram a correlação entre as variáveis da seguinte forma: variáveis próximas refletem uma correlação altamente positiva, enquanto que variáveis diagonalmente opostas apresentam uma correlação muito negativa. A Componente Principal que melhor as diferencia também pode ser indicada por este tipo de gráfico.

No caso da Figura 8, a maior separação da cocaína e da cafeína ocorre na PC1, a qual explica 65% da variância total do sistema. A lidocaína não é bem explicada nem pela PC1, nem pela PC2, estando mais próxima da origem e a maior separação da cocaína e da fenacetina ocorre na PC2, que explica 23% da variância total do sistema.

Entretanto, este tipo de gráfico é pouco informativo conforme se aumenta o número de variáveis, pois pode ser muito complexo e os valores de *loadings* podem se sobrepor.

### 3.6 Pré-Processamento de dados

Os tipos de pré-processamento de dados são vários (Tabela 2) (ESBENSEN, 2018; GEMPERLINE, 2006), e dependem de cada situação, não podendo ser escolhidos aleatoriamente, pois um pré-processamento usado erroneamente é capaz de destruir a estrutura original dos dados. O principal objetivo de seu uso é remover fontes indesejáveis de variação advindas de ruídos e efeitos externos, de modo que a estrutura latente dos dados seja revelada e modelada (ESBENSEN, 2018).

Os principais tipos de pré-processamento são os gerais, para variáveis discretas, e os espectroscópicos, que são direcionados para dados espectroscópicos por estes apresentarem

algumas peculiaridades. As técnicas gerais são usualmente o primeiro passo a ser dado, seguido pela avaliação do pré-processamento adequado aos dados obtidos.

Nas análises espectroscópicas, há duas principais fontes de variação, que exigem o uso de algum pré-processamento mais específico: os efeitos aditivos (gerados pelas diferenças de empacotamento e densidade da amostra) e os efeitos multiplicativos (causados pelo fenômeno do espalhamento de luz).

Tabela 2 – Técnicas de pré-processamento mais utilizadas. Há dois tipos principais de pré-processamento: geral, para variáveis discretas e espectroscópico, para variáveis espectroscópicas.

<b>Pré-Processamento Geral</b>			
<b>Nome traduzido</b>	<b>Nome original</b>	<b>Sigla</b>	<b>Função</b>
<b>Centralização pela média</b>	<i>Mean centering</i>	--	Centralizar os dados no objeto médio ou num ponto que corresponda à média dos objetos, subtraindo a matriz pelo valor médio da coluna correspondente.
<b>Escala por variância</b>	<i>Variance scaling</i>	--	Colocar as variáveis numa mesma escala de variância, através da divisão de cada uma pelo desvio padrão correspondente.
<b>Auto escala</b>	<i>Auto scaling</i>	--	Unir os benefícios da centralização pela média aos do escala por variância. Assim, pequenas mudanças em uma variável podem ser comparadas com outra.
<b>Pré-Processamento Espectroscópico</b>			
<b>Suavização</b>	<i>Smoothing</i>	--	Reduzir o ruído sem reduzir as variáveis.
<b>Correção de linha-base</b>	<i>Baseline correction</i>	--	Remover os efeitos aditivos.
<b>Derivadas</b>	<i>Derivatives</i>	--	Suavizar e remover os efeitos aditivos.
<b>Variável Normal Padrão</b>	<i>Standard Normal Variate</i>	SNV	Corrigir os efeitos multiplicativos.
<b>Correção do Efeito Multiplicativo por Espalhamento</b>	<i>Multiplicative Scatter Correction</i>	MSC	Remover os efeitos aditivos e multiplicativos.

#### 4 ESTRATÉGIA EXPERIMENTAL

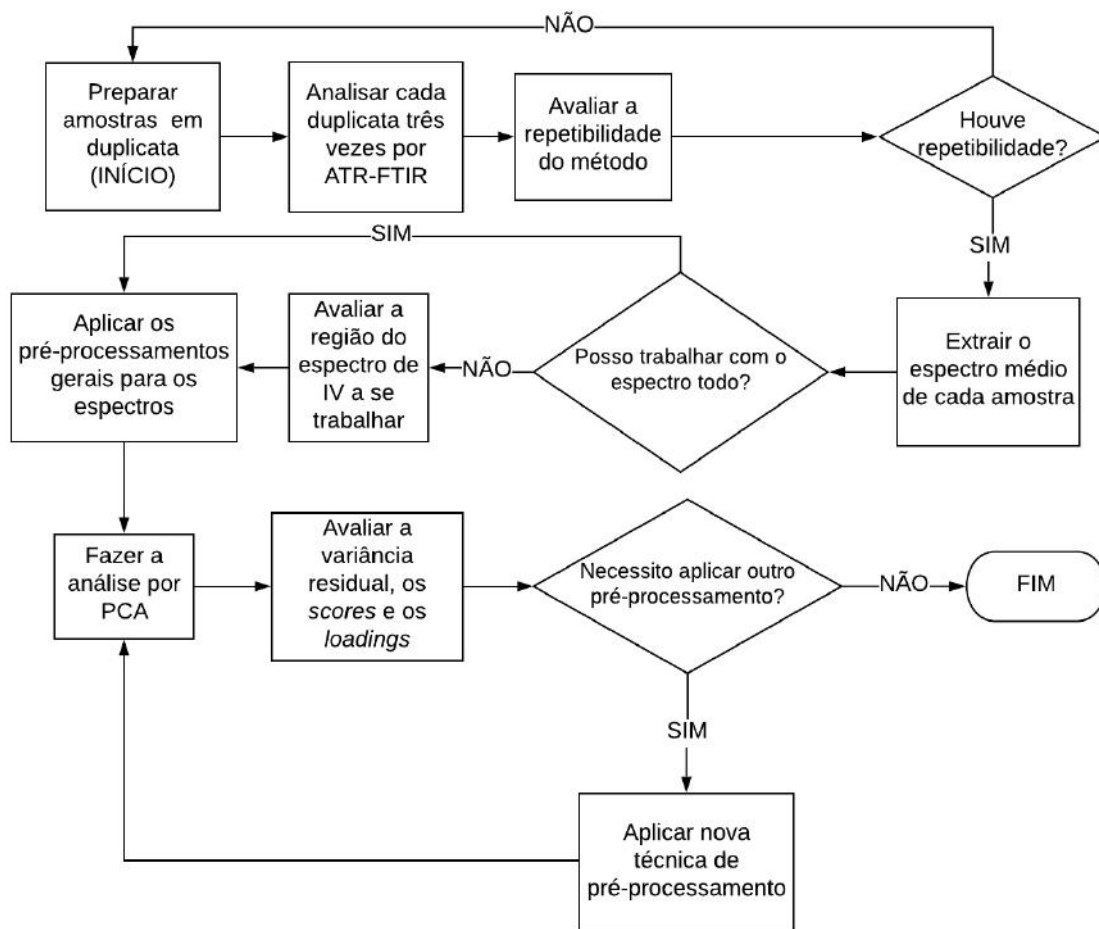


Figura 9 – Fluxograma explicativo da estratégia experimental adotada no presente trabalho. O início e o fim estão destacados, para melhor esclarecimento. O planejamento das amostras binárias e ternárias de cocaína, lidocaína e cafeína buscou variar a concentração de cada um dos três componentes em cerca de 10%, de modo que abrangesse o intervalo de 0-100% para todos os três compostos.

## 5 METODOLOGIA

### 5.1 Avaliação da repetibilidade do método e estabilidade do equipamento

#### 5.1.1 Preparo das amostras

As misturas binárias e ternárias foram preparadas através de pesagem de cada componente da amostra em tubo plástico cônico dotado de tampa acoplada (*eppendorf*), previamente tarado em balança microanalítica Mettler Toledo XPE (Figura 10), com resolução de 10 µg, de modo que a massa total da mistura perfizesse o valor de  $(100,0 \pm 0,1)$  mg. A homogeneização foi feita dentro do tubo plástico, com instrumento disponível no laboratório. Cada mistura foi feita em duplicata.

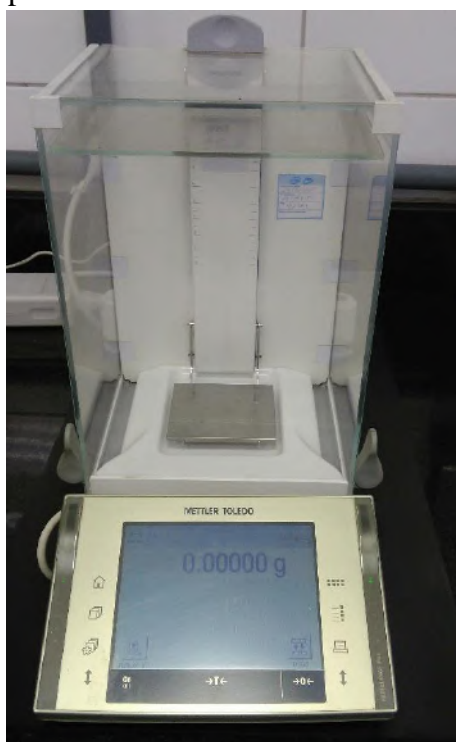


Figura 10 – Balança microanalítica Mettler Toledo XPE, com resolução de 10 µg, utilizada para pesagem das amostras preparadas.

#### 5.1.2 Análise por ATR-FTIR

As amostras foram submetidas à análise por infravermelho com transformada de Fourier num espectrômetro Nicolet 380 FTIR (Figura 11) equipado com acessório de refletância total atenuada (ATR) Smart Orbit, usando um cristal de diamante como elemento de reflexão interna. Cada duplicata foi medida três vezes, a partir de diferentes pontos do *eppendorf* (Figura 12).

As medidas foram coletadas no modo absorvância. O *background* espectral foi coletado antes de cada mistura diferente. Os espectros foram coletados na região de  $4000\text{--}400\text{ cm}^{-1}$ , com 32 scans e resolução de  $4\text{ cm}^{-1}$ . Os dados de absorvância foram advindos do programa *EZ Omnic*, utilizado pelo espectrômetro de infravermelho para a geração dos espectros. Os picos de maior intensidade foram assinalados de acordo com o critério estabelecido no programa.



Figura 11 – Espectrofotômetro Nicolet 380 FT-IR equipado com acessório de refletância total atenuada (ATR) Smart Orbit, usando um cristal de diamante como elemento de reflexão interna.

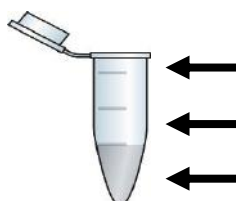


Figura 12 - Locais do eppendorf a partir dos quais se tiraram alíquotas de cada duplicata para que fossem analisadas por ATR-FTIR.

## 5.2 Avaliação da região do espectro a se trabalhar

A região do espectro a se trabalhar foi aquela que apresentou, visualmente, maior informação química dos compostos.

## 5.3 Aplicação dos pré-processamentos gerais para os espectros

A matriz  $n \times p$  foi organizada de acordo com os princípios algébricos da Análise por Componentes Principais (PCA). Foi feita a centralização pela média dos dados e o resultado foi lançado no programa de estatística, que fez o escalamento por variância dos dados.

#### **5.4 Análise por Componentes Principais preliminar**

A Análise por Componentes Principais dos dados foi feita usando-se o programa de estatística. Foram avaliadas as variâncias residuais de cada PC a fim de que se obtivesse o número ótimo de PCs, os gráficos de *scores* e *loadings* e fez-se necessário um novo pré-processamento.

#### **5.5 Novo pré-processamento e Análise por Componentes Principais**

O novo pré-processamento adotado foi o de primeira derivada para a matriz  $n \times p$ . A nova matriz transformada foi submetida aos mesmos pré-processamentos gerais e a uma nova Análise por Componentes Principais.

## 6 RESULTADOS E DISCUSSÃO

### 6.1 Avaliação da repetibilidade do método e da estabilidade do equipamento

A coleta de alíquotas em diferentes pontos do *eppendorf* buscou verificar a homogeneização correta de cada mistura. Em todas as misturas binárias e ternárias, avaliou-se se a homogeneização estava adequada.

Como todos os 6 espectros foram avaliados para cada amostra, individualmente, e não foi percebida visualmente nenhuma tendência ou diferença significativa, foi possível perceber que a homogeneização foi adequada, o que permitiu considerar as duas amostras de uma mesma composição como duplicatas. A Figura 13 apresenta o exemplo das seis réplicas de uma amostra. O não deslocamento das bandas também permitiu que se verificasse uma repetibilidade satisfatória nos espectros.

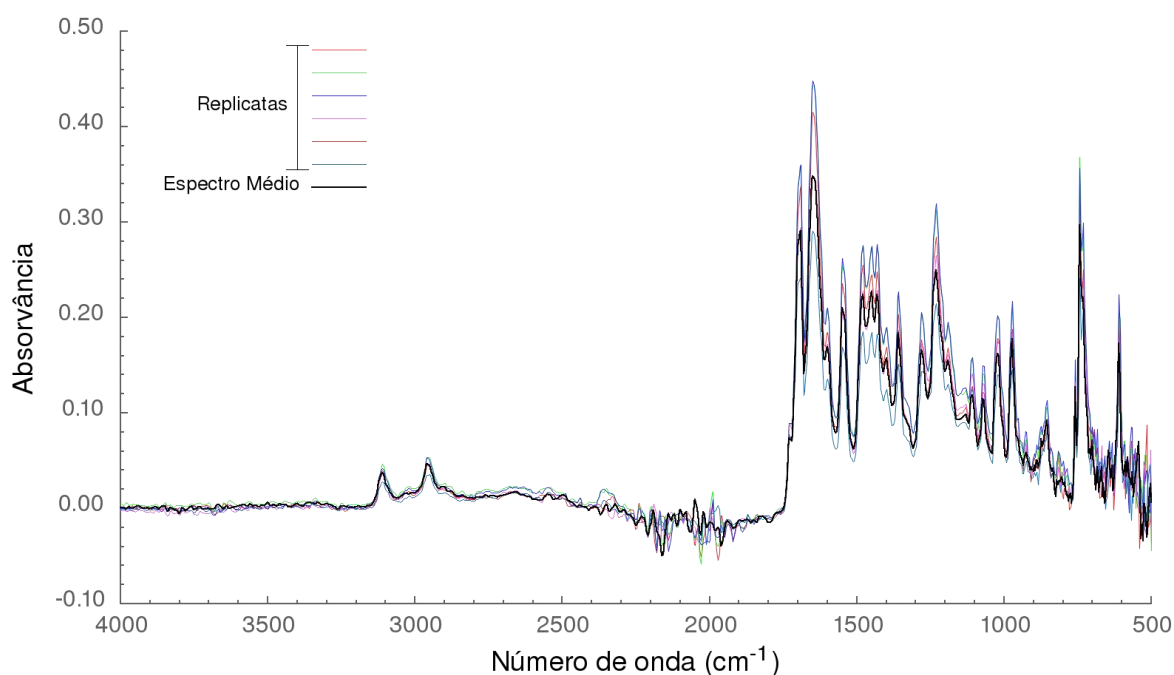


Figura 13 - Espectros de todas as seis réplicas de uma das amostras e seu espectro médio correspondente.

A variação dos valores de absorvância dos espectros de uma mesma amostra precisa ser considerada, indubitavelmente, para uma análise quantitativa. Esta variação pode ser atribuída à diferença de caminho ótico percorrido pela luz nas amostras, ainda que de mesma composição,

gerando efeitos aditivos e multiplicativos. Contudo, para a PCA, por ser de caráter exploratório e visual, trabalhar-se-á com os espectros médios.

## 6.2 Avaliação da região do espectro a se trabalhar

Os primeiros espectros analisados foram os correspondentes aos três compostos puros (Figura 14). As duas regiões em destaque ( $4000\text{--}3500\text{ cm}^{-1}$ ) e ( $2300\text{--}1850\text{ cm}^{-1}$ ) são passíveis de gerar grandes problemas no método de PCA, tendo em vista que estas regiões apresentam grande dispersão, sem, contudo, gerar nenhuma informação química relevante. Adicionalmente, a região de  $2100\text{--}1900\text{ cm}^{-1}$  corresponde à de absorção do elemento de reflexão total interna, não podendo, por isso, ser utilizada.

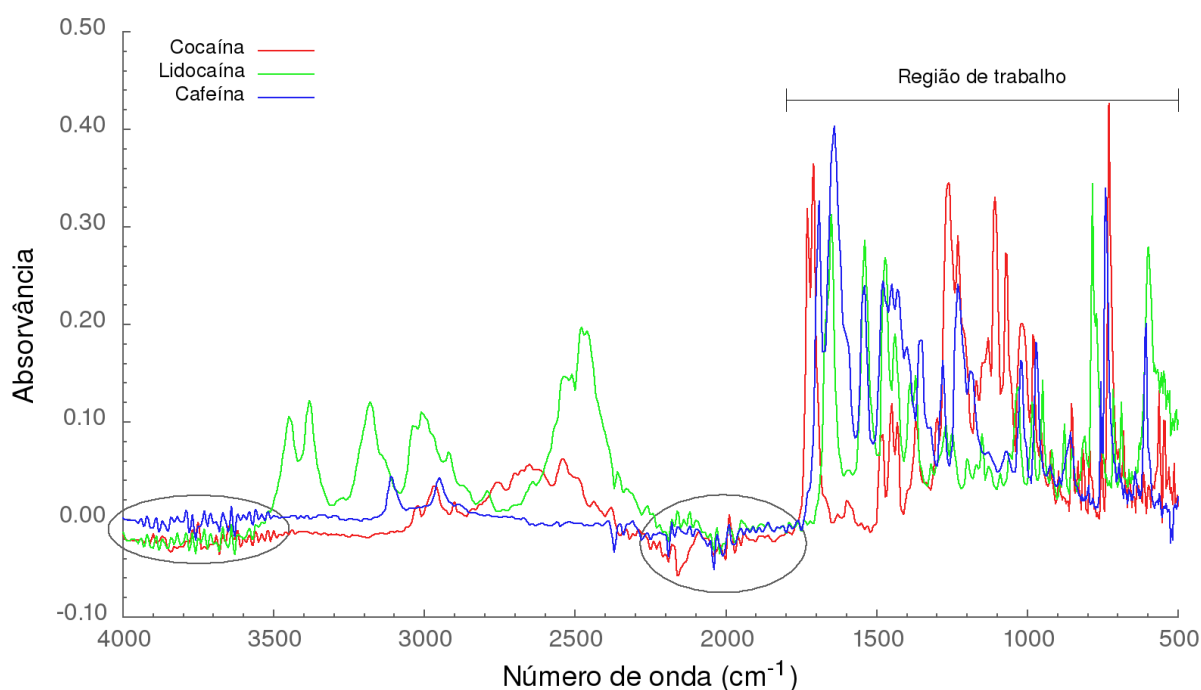


Figura 14 – Espectros de infravermelho de cocaína, na forma de cloridrato, cloridrato de lidocaína e cafeína puras, na região de  $4000\text{--}500\text{ cm}^{-1}$ , coletados pelo espectrofotômetro Nicolet 380 ATR-FTIR. As regiões em destaque foram as primeiras a serem excluídas, devido à alta variação aleatória presente (ruído).

A região entre essas duas ( $3500\text{--}2300\text{ cm}^{-1}$ ) apresenta sinais de intensidade bem baixa para a cocaína e a cafeína; portanto, preferiu-se trabalhar apenas na região de  $1850\text{--}500\text{ cm}^{-1}$ . A mesma escolha foi feita por Eliaerts *et al.* (2017), Marcelo *et al.* (2015) e Penido *et al.* (2017). A análise apenas dessa região também apresentou uma aplicação direta na análise de amostras em campo, através de um espectrômetro de infravermelho portátil (HESPANHOL, 2019).



### 6.2.1 Interpretação química dos espectros dos compostos puros

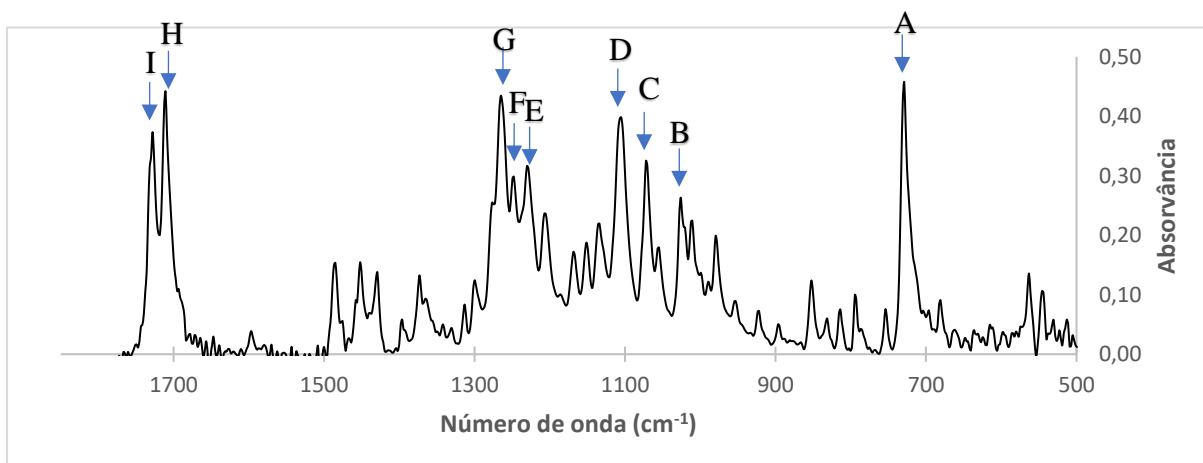


Figura 15 – Espectro de infravermelho na região de trabalho (1850-500  $\text{cm}^{-1}$ ) do cloridrato de cocaína. As letras correspondem às bandas assinaladas na Tabela 3.

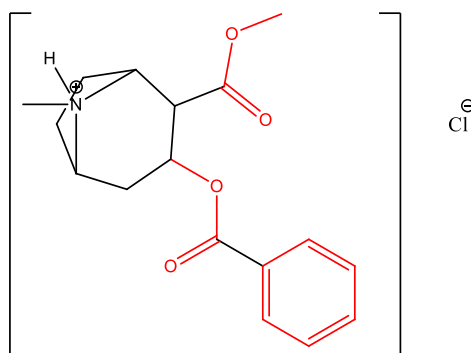


Figura 16 - Estrutura química da cocaína na forma de cloridrato (Fonte: SOUZA, 2016)

Tabela 3 - Interpretação dos fenômenos químicos vibracionais que resultam nas bandas de maior intensidade (indicadas na Figura 15) do espectro da cocaína na forma de cloridrato na região de 1850-500  $\text{cm}^{-1}$  (Fontes: COATES, 2006; GROBÉRIO, 2014; MAHARAJ, 2008; RODRIGUES, 2013).

Interpretação das bandas do espectro de Cocaína		
Região	Bandas principais ( $\text{cm}^{-1}$ )	Interpretação química
<b>A</b>	729	Deformação fora do plano do anel aromático Estiramento do anel benzílico monossustituído
<b>B</b>	1026	
<b>C</b>	1070	
<b>D</b>	1107	Estiramento da ligação C - O
<b>E</b>	1230	
<b>F</b>	1250	
<b>G</b>	1265	
<b>H</b>	1712	Estiramento dos dois grupos de carbonila de éster (C=O)
<b>I</b>	1728	

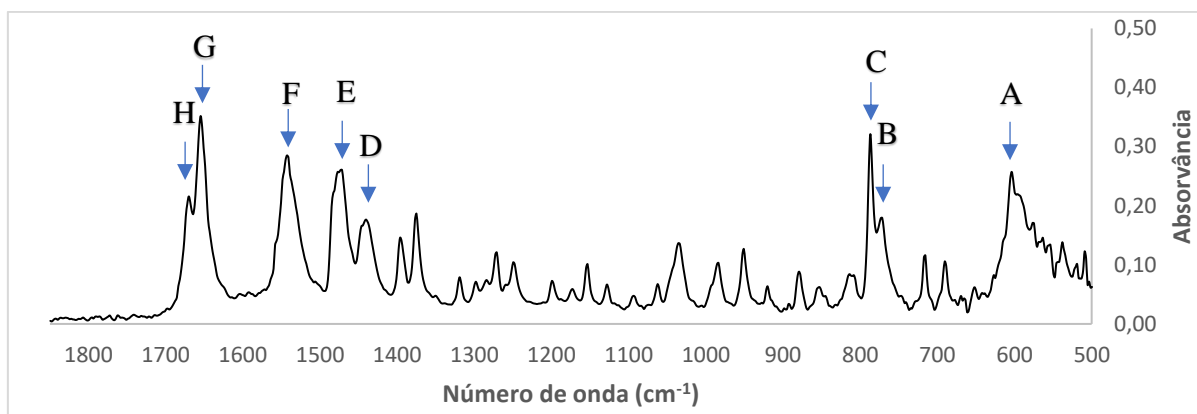


Figura 17 - Espectro de infravermelho na região de trabalho (1850-500  $\text{cm}^{-1}$ ) do cloridrato de lidocaína. As letras correspondem às bandas assinaladas na Tabela 4.

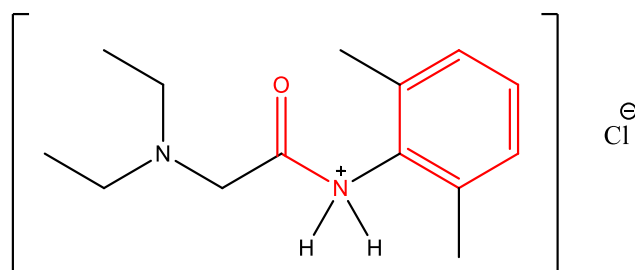


Figura 18 - Estrutura química da lidocaína na forma de cloridrato (Fonte: SOUZA, 2016)

Tabela 4 - Interpretação dos fenômenos químicos vibracionais que resultam nas bandas de maior intensidade (indicadas na Figura 17) do espectro da lidocaína na forma de cloridrato na região de 1850-500  $\text{cm}^{-1}$  (Fontes: COATES, 2006; RODRIGUES, 2013; SILVERSTEIN, 2005).

Interpretação das bandas do espectro do cloridrato de Lidocaína		
Região	Bandas principais ( $\text{cm}^{-1}$ )	Interpretação química
A	606	Bandas de anel aromático
B	773	Deformação angular do N-H fora do plano
C	787	
D	1475	Bandas de anel aromático
E	1477	
F	1541	Estiramento C=C do anel aromático Distorção angular N-H no plano
G	1654	Estiramento C=O de amida
H	1670	Estiramento N-H de amida

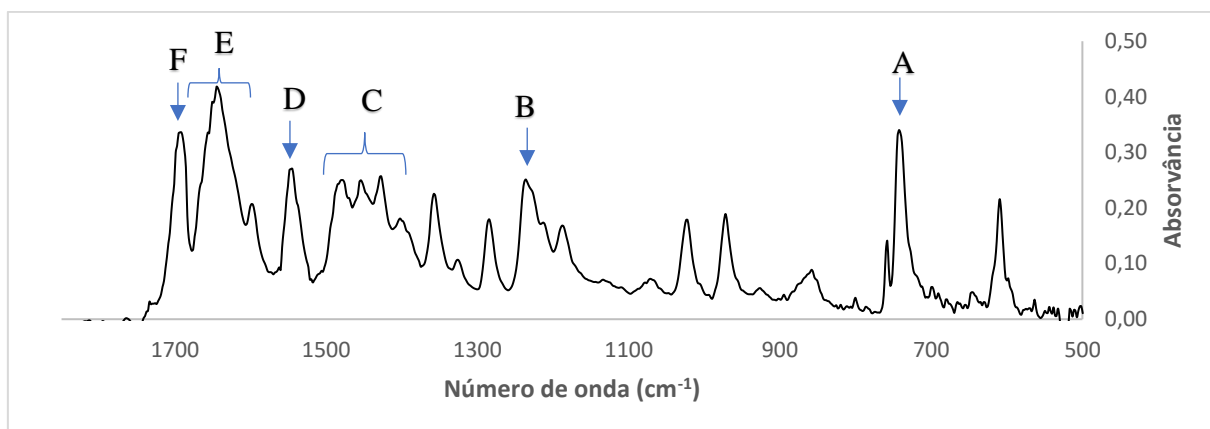


Figura 19 - Espectro de infravermelho na região de trabalho (1850-500  $\text{cm}^{-1}$ ) da cafeína. As letras correspondem às bandas assinaladas na Tabela 5.

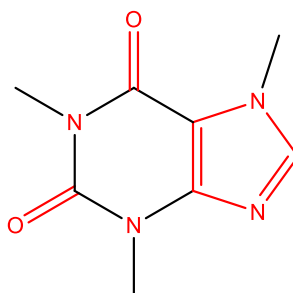


Figura 20 - Estrutura química da cafeína (Fonte: SOUZA, 2016)

Tabela 5 - Interpretação dos fenômenos químicos vibracionais que resultam nas bandas de maior intensidade (indicadas na Figura 19) do espectro da lidocaína na forma de cloridrato na região de 1850-500  $\text{cm}^{-1}$  (Fontes: KATON, 1990; RAMASAMY, 2015; RODRIGUES, 2013).

Interpretação das bandas do espectro da cafeína		
Região	Bandas principais ( $\text{cm}^{-1}$ )	Interpretação química
<b>A</b>	742	Distorção angular C-H fora do plano
<b>B</b>	1236	Distorção angular N-H no plano
<b>C</b>	1429	Estiramento C-N do imidazol
	1456	
	1483	
	1486	
<b>D</b>	1549	Vibração C=O
		Estiramento C=N
<b>E</b>	1645	Estiramentos C=O
	1651	
	1657	
<b>F</b>	1695	Vibração C=O

Os espectros de compostos puros podem ser interpretados através da consulta à vasta literatura da espectroscopia no infravermelho. Entretanto, quando as mesmas substâncias estão presentes em misturas, como geralmente ocorre com a *cocaína de rua*<sup>6</sup>, esse trabalho já não é mais possível. A Figura 21 representa os espectros médios de todas as amostras preparadas neste trabalho.

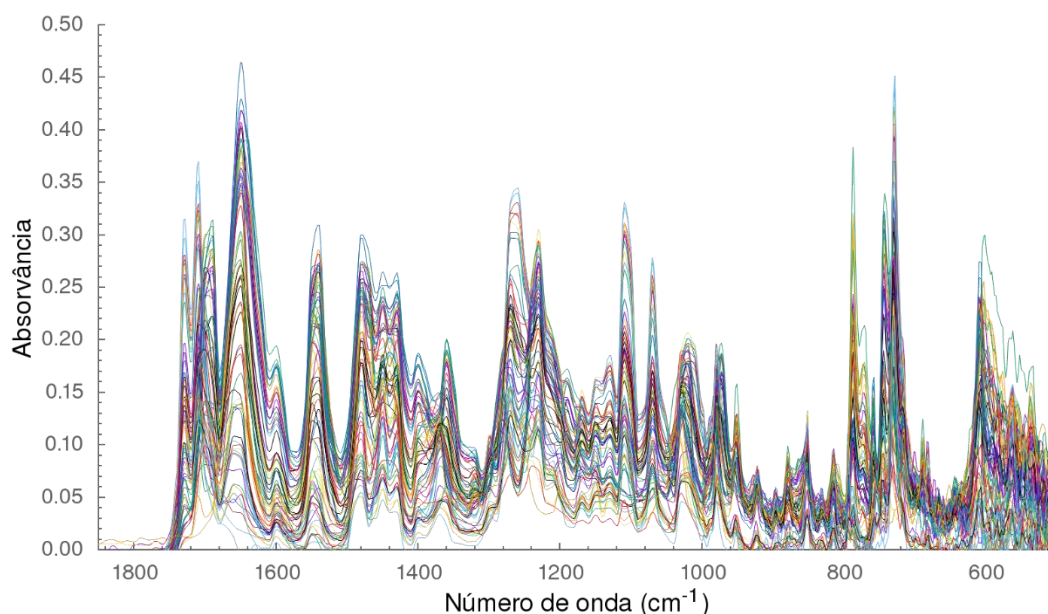


Figura 21 – Espectros de todas as amostras analisadas, coletados pelo espectrofotômetro Nicolet 380 FT-IR, na região de 1850-500  $\text{cm}^{-1}$ .

Conforme percebido, a introdução de um ou mais componentes a um composto puro aumenta significativamente a complexidade do espectro, fazendo com que ele perca tanto seu aspecto qualitativo quanto quantitativo. Entretanto, é possível recuperar suas informações de ambos os quesitos com as ferramentas da Análise Multivariada.

### 6.3 Aplicação dos pré-processamentos gerais para os espectros

A matriz  $n \times p$  foi organizada, de modo que as linhas fossem os objetos (amostras) e as colunas, as variáveis (números de onda). Em seguida, a centralização pela média e o escalamento por variância, que compõem o auto escalamento.

<sup>6</sup> O termo é usualmente utilizado para descrever o tipo de cocaína comumente comercializado nas ruas.

## 6.4 Análise por Componentes Principais preliminar

Após o pré-processamento, a matriz foi submetida à Análise por Componentes Principais. O primeiro ponto a ser avaliado foi o número ótimo de PCs.

### 6.4.1 Análise da variância residual

Através da regra empírica, deduz-se que o número ótimo de PCs a ser utilizada é igual a 3, pois neste valor, há uma significativa queda da inclinação da curva de variância residual relativa (Figura 22). É possível ver também que a partir do quarto componente principal, a variância residual relativa assume valor praticamente constante.

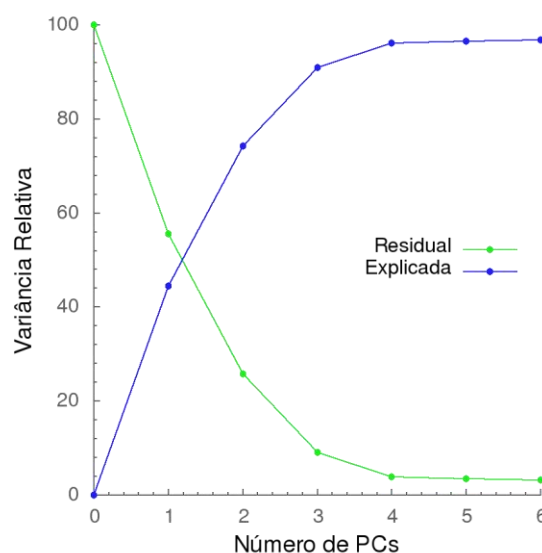


Figura 22 – Variâncias residual e explicada da matriz de dados à medida que se processa a Análise por Componentes Principais, após a aplicação de pré-processamentos gerais. As variâncias residual e explicada, ao serem somadas, perfazem um valor de 100%. Elas são relativas ao valor residual total absoluto, extraído da matriz de resíduos (E).

A variância também pode ser expressa em termos do quanto é explicado pelo modelo, atentando-se para o fato de que a soma das variâncias residual e explicada é igual a 100%. Assim, verifica-se que a PC1 explica 44,52% da variância total do sistema; a PC2, 29,70%; e a PC3, 16,76%, totalizando 90,98% (Tabela 6).

Tabela 6 – Variância explicada por cada PC calculado e a cumulativa correspondente para o sistema, após aplicação dos métodos de pré-processamentos gerais. Em destaque, o número ótimo de PCs definido para regra empírica.

Componente Principal	Explicação da variância do sistema (%)	Explicação cumulativa da variância do sistema (%)
<b>0</b>	0	0
<b>1</b>	44,52	44,52
<b>2</b>	29,70	74,22
<b>3</b>	<b>16,76</b>	<b>90,98</b>
<b>4</b>	5,21	96,18
<b>5</b>	0,33	96,51
<b>6</b>	0,24	96,76

De forma complementar, pode-se avaliar a eficiência da PCA através da evolução da queda da variância absoluta da matriz antes e depois de se calcularem os PCs (Figura 23). A variância residual absoluta cai para quase metade já na PC1 e para 9,02% do valor original ao longo de 3 PCs, corroborando que este valor é bastante satisfatório.

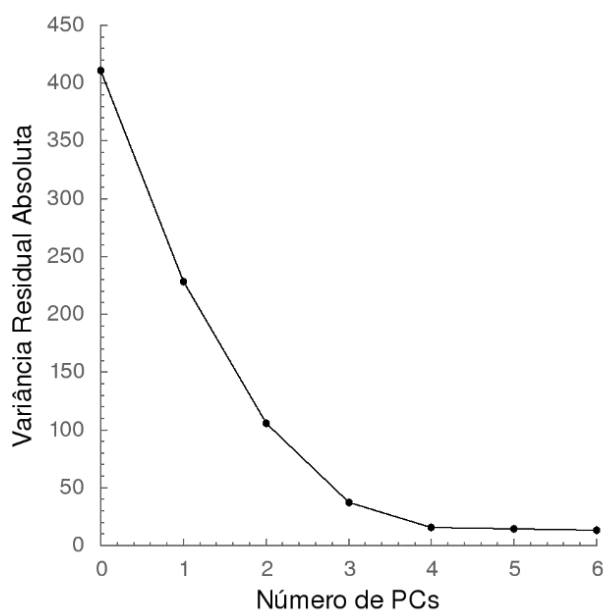


Figura 23 - Variância residual absoluta da matriz de dados e a evolução do seu decréscimo conforme são calculadas as Componentes Principais. A PC1 faz com que a variância residual caia para 55,48% do valor original; a PC2, para 25,78% e a PC3, para 9,02%.

#### 6.4.2 Análise dos scores

No primeiro gráfico de scores, PC1 vs. PC2, o único grupo de amostras que se destacou foi o grupo de misturas binárias de lidocaína e cafeína, incluindo os seus compostos puros (Figura 24), conforme destacado, separado pela PC2. O grupo de misturas ternárias em destaque pela PC1 não apresenta nenhum padrão de comportamento.

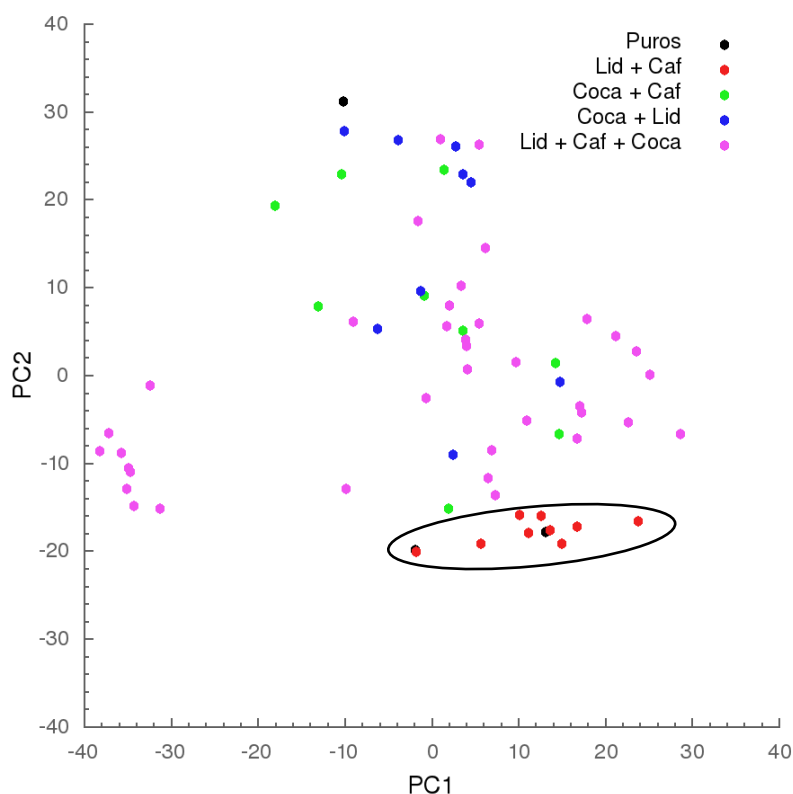


Figura 24 - Gráfico de *scores* PC1 vs. PC2, que destaca o grupo de misturas binárias, e seus padrões correspondentes, de lidocaína e cafeína ao longo da PC2. A PC1 separou um grupo de amostras ternárias que não apresentam, porém, nenhuma característica em comum.

No segundo gráfico, PC1 vs. PC3, foi apresentado padrão semelhante (Figura 25). Na região em destaque, a PC3 separou misturas binárias de cocaína e cafeína, com seus respectivos componentes puros. Logo, é possível perceber que explicar uma maior variância do sistema, como no caso da PC1, não necessariamente implica em um resultado que faça sentido quimicamente.

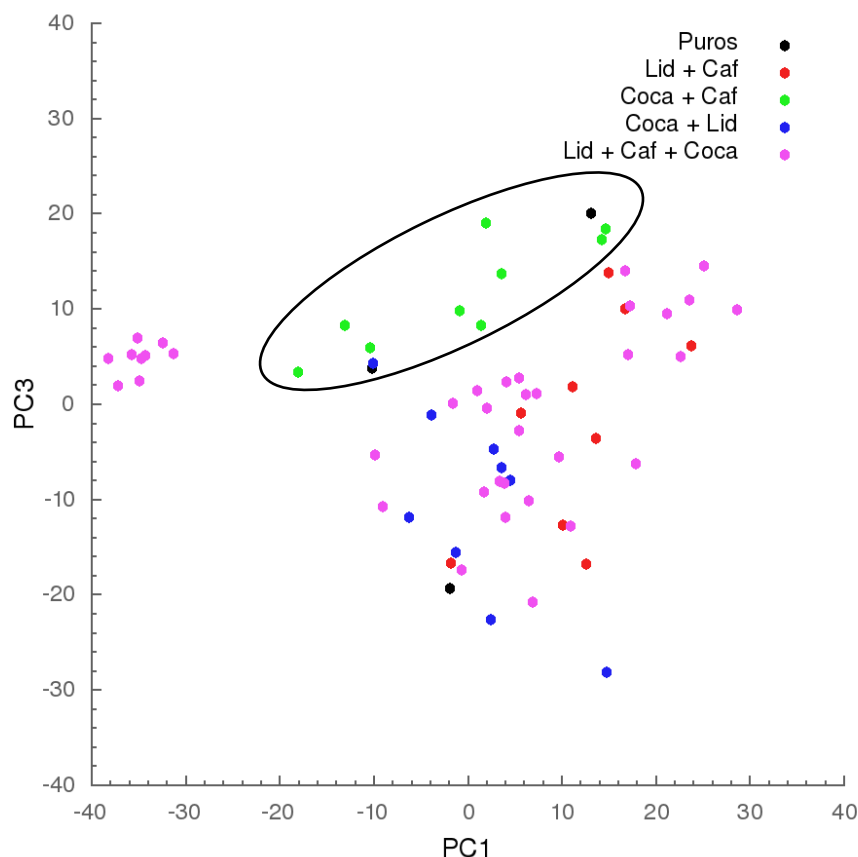


Figura 25 - Gráfico de *scores* PC1 vs. PC3, que destaca o grupo de misturas binárias, e seus padrões correspondentes, de cocaína e cafeína ao longo da PC3.

Ao se analisar, porém, o gráfico de *scores* PC3 vs. PC2 (Figura 26), percebe-se que os componentes puros estão diagonalmente opostos entre si e, quando ligados, formam os vértices de um triângulo, o que é um indicativo de uma possível classificação de acordo com a composição das amostras, conforme um diagrama de três componentes.

Como a descrição dos dados analisados pode ser feita tanto a partir de uma análise exploratória de dados como de uma modelagem empírica da estrutura dos dados (ESBENSEN, 2018), é possível classificar as amostras de acordo com algum critério coerente que seja adotado pelo analista.

Neste caso, o critério adotado seria classificar as amostras como binárias ou ternárias de acordo com a sua posição no gráfico de *scores* – seria esperado que as amostras binárias se encontrassem ao longo das arestas do triângulo, e as ternárias, no interior do triângulo. Como isso não foi observado, foi necessário um novo pré-processamento.



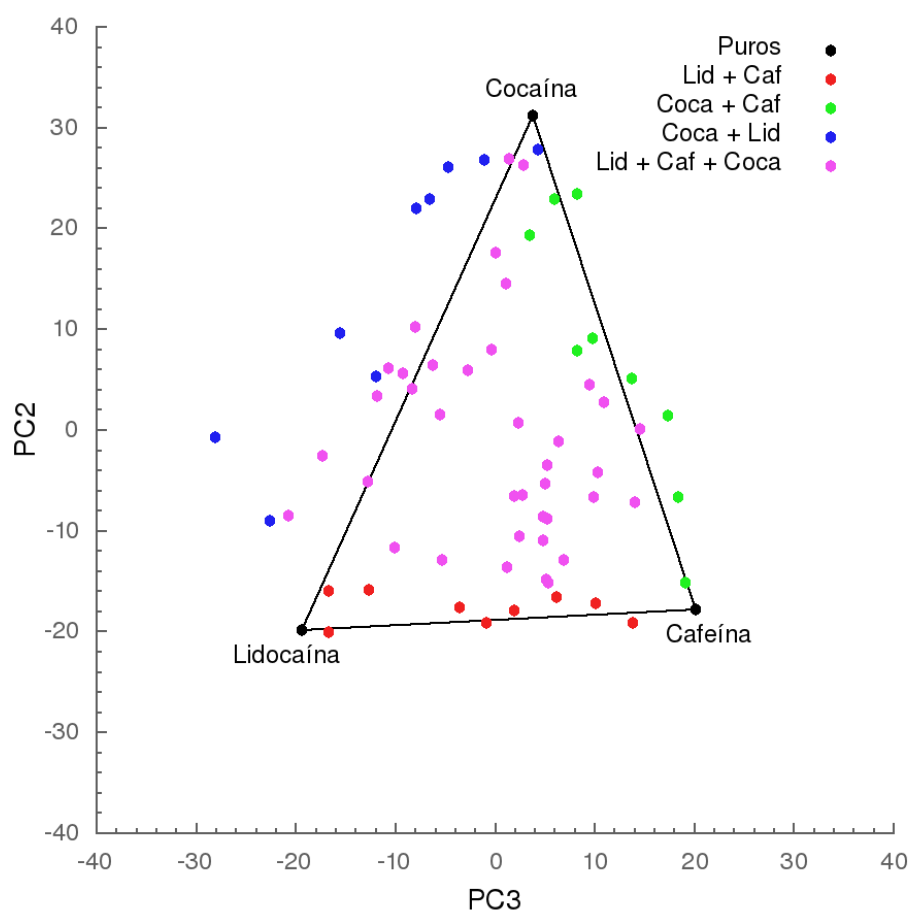


Figura 26 – Gráfico de scores PC3 vs. PC2, indicando uma relação entre as amostras similar à de um diagrama de três componentes, tornando a possibilidade de classificação por composição promissora

#### 6.4.3 *Análise dos loadings*

Com a alta quantidade de variáveis, o gráfico 2-D dos loadings se torna pouco útil pelo excesso de informações e baixa possibilidade de correlações, conforme exposto pela Figura 27. A avaliação dos loadings correspondentes a cada PC é, portanto, mais pertinente.

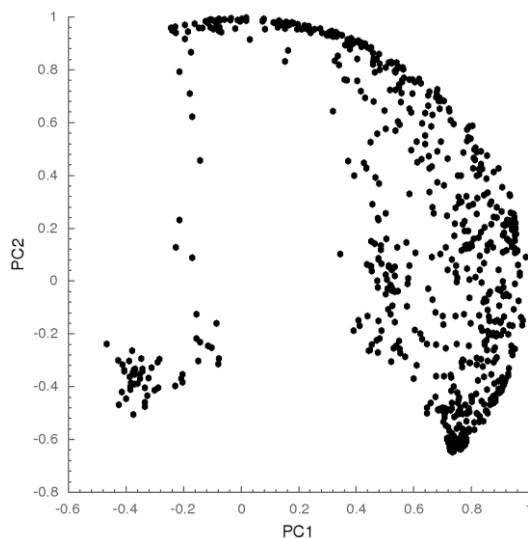


Figura 27 – Gráfico de *loadings* PC1 vs. PC2.

A partir da Figura 28, é possível observar que as regiões correspondendo a 730, 1070-1100 e 1267  $\text{cm}^{-1}$  não contribuem para a PC1, pois o seu valor de *loading* é em torno de zero. As regiões de 730 e 1267  $\text{cm}^{-1}$  são de sobreposição entre bandas dos três compostos, enquanto que a região a partir de 1708  $\text{cm}^{-1}$  já se aproxima dos ruídos, onde a absorvância dos compostos está retornando à linha base.

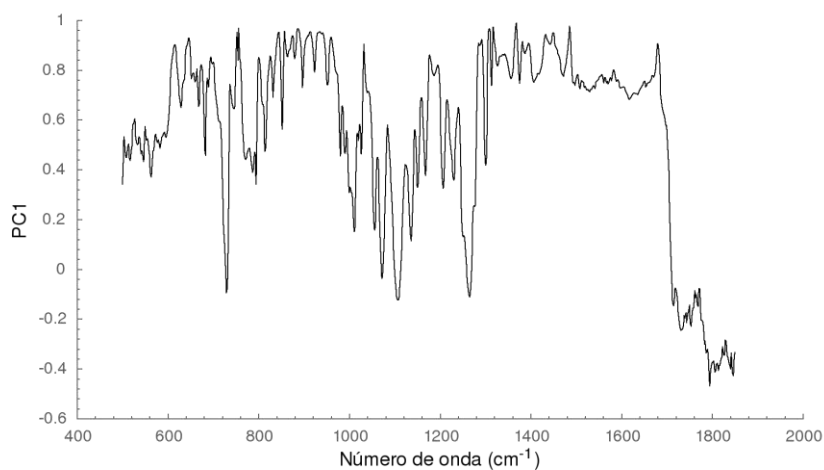


Figura 28 – Gráfico de *loadings* para a primeira Componente Principal (PC1)

Pelo gráfico de loadings em função da PC2, há mais regiões do espectro que contribuem para a PC2 (com valor relativo próximo de 1), conforme Figura 29, sendo um possível motivo pelo qual a PC2 permite observar outras informações dos espectros, conseguindo, assim, ver uma classe de misturas. O mesmo é verificado para a PC3 (Figura 30). Embora tenha sido

possível extrair alguma informação para esse conjunto de dados, elas ainda não são suficientes para a classificação desejada.

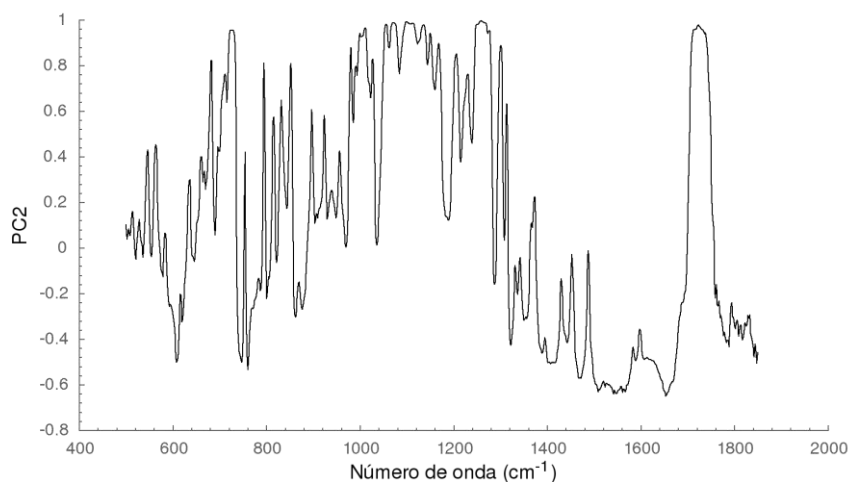


Figura 29 - Gráfico de *loadings* para a segunda Componente Principal (PC2)

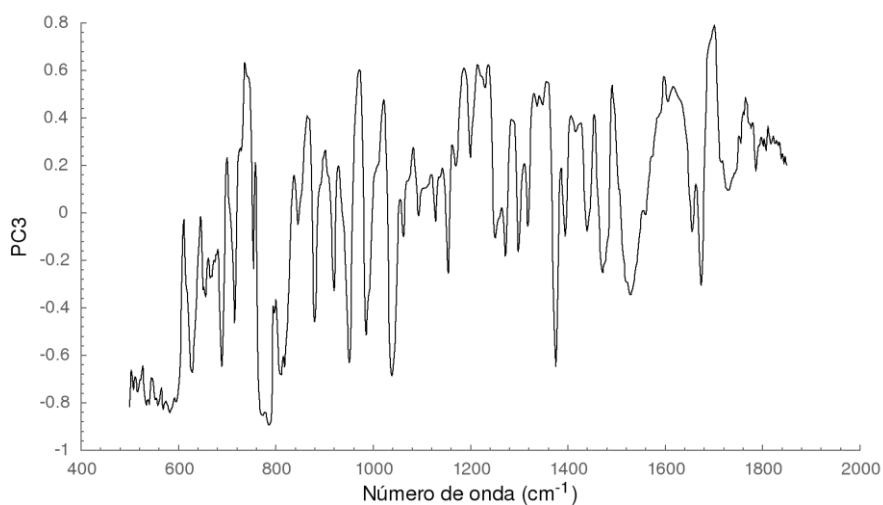


Figura 30 - Gráfico de *loadings* para a terceira Componente Principal (PC3)

## 6.5 Novo pré-processamento e Análise por Componentes Principais

### 6.5.1 Características da técnica de pré-processamento escolhida

Dentre as técnicas espectroscópicas de pré-processamento, foi escolhida a primeira derivada. A transformação por primeira derivada é calculada através da diferença entre o segundo ponto do espectro e o primeiro, dividida por uma constante de centralização (ESBENSEN, 2018). Portanto, a derivada centra o espectro em torno do zero, algo que é muito

importante, pois flutuações nas medições fizeram com que o mínimo de muitos espectros de absorvância medidos estivesse no zero.

Entretanto, é uma técnica que pode amplificar ruídos; por isso, é importante que toda a região que seja apenas ou majoritariamente ruidosa seja retirada previamente.

Com estes cuidados, as derivadas dos dados da matriz original foram calculadas e a matriz resultante foi novamente submetida à PCA.

### 6.5.2 *Análise da variância residual*

Retomando ao primeiro passo pós-PCA, foi avaliado o número ótimo de PCs a partir do gráfico de variância residual relativa. A partir da Figura 31, percebe-se que o número ótimo de PCs, pela regra empírica, é de 2, onde ocorre a maior diferença de inclinação da curva da Variância Residual Relativa e a partir de 3 PCs, esta variância não se altera consideravelmente.

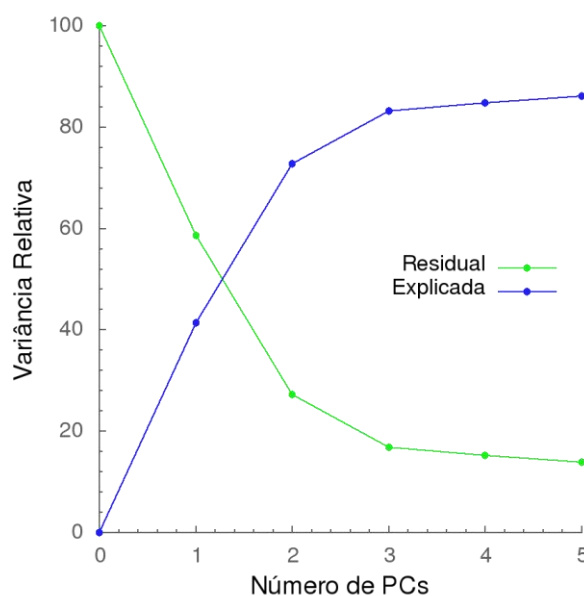


Figura 31 - Variâncias residual e explicada da matriz de dados à medida que se processa a Análise por Componentes Principais, após a aplicação de pré-processamentos gerais e da transformação por primeira derivada.

A primeira Componente Principal explica 41,39% da variância do sistema e a PC2, 31,43%, totalizando 72,82% da variância total explicada (Tabela 7). Isto representa uma perda de quase 20% da variância total explicada pelo modelo, porém é importante lembrar que a transformada por derivada amplifica ruídos, sendo possível que essa diminuição não gere nenhuma perda significativa de variância importante para o modelo químico a ser construído.

Tabela 7 - Variância explicada por cada PC calculado e a cumulativa correspondente para o sistema, após aplicação dos métodos de pré-processamentos gerais e por primeira derivada. Em destaque, o número ótimo de PCs definido para regra empírica.

Componente Principal	Explicação da variância do sistema (%)	Explicação cumulativa da variância do sistema (%)
0	0	0
1	41,39	41,39
<b>2</b>	<b>31,43</b>	<b>72,82</b>
3	10,35	83,17
4	1,64	84,81
5	1,35	86,16

### 6.5.3 Análise de scores

A análise de *scores* das duas primeiras componentes principais precisa ser suficiente para que se estabeleça o modelo classificatório desejado, se de fato o número ótimo de PCs for igual a 2, o que foi inferido através do gráfico de *scores* PC1 vs. PC2 (Figura 32).

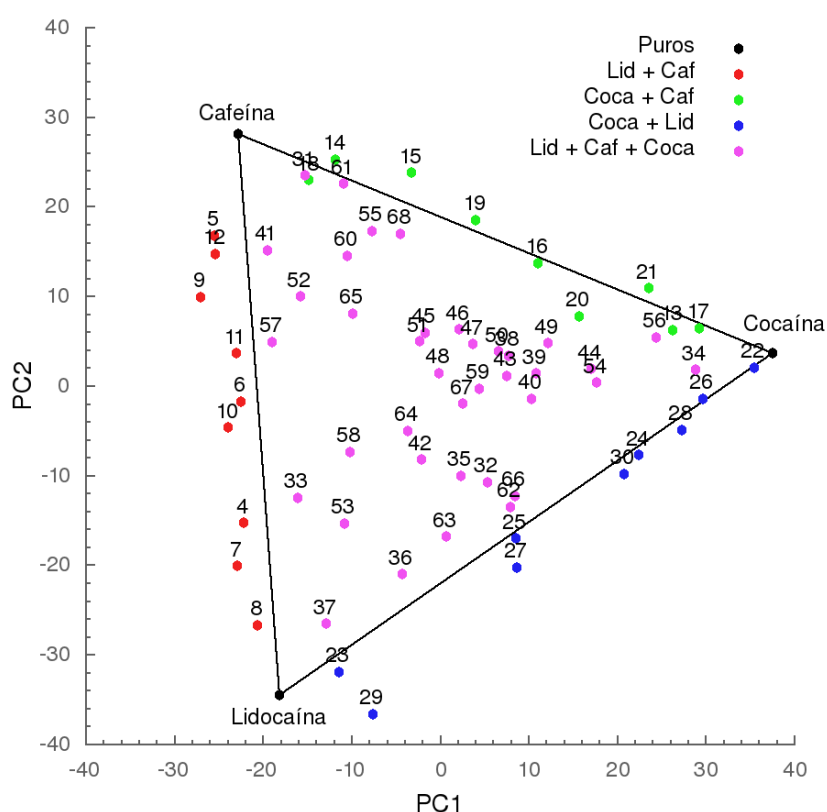


Figura 32 - Gráfico de *scores* PC1 vs. PC2, para os dados pré-processados por: auto escalamento e transformada por primeira derivada. É nítida a possível classificação por composição, de acordo com o diagrama de três componentes.

Foi vista a relação entre as amostras de acordo com a sua composição, sendo possível classifica-las<sup>7</sup> dessa forma. É possível ainda, através de uma avaliação empírica, detectar quais não satisfazem o critério estabelecido a fim de se detectarem as limitações do modelo desenvolvido.

Considerando o critério de que todas as amostras que estejam dentro do triângulo sejam consideradas ternárias e as que estiverem ao longo das arestas do triângulo ou fora, consideradas binárias, verifica-se que as amostras foram classificadas conforme expresso na Tabela 8.

Tabela 8 – Resultado da classificação das amostras, a partir do gráfico de *scores* PC2 vs. PC1 dos dados pré-processados por auto escalamento e transformação por primeira derivada. As amostras que apresentaram uma classificação errada estão em destaque.

<b>Amostra</b>	<b>Mistura</b>	<b>Classificação</b>	<b>Resultado</b>
<b>1</b>	COC	-	-
<b>2</b>	LID	-	-
<b>3</b>	CAF	-	-
<b>4</b>	LID + CAF	LID + CAF	Acerto
<b>5</b>	LID + CAF	LID + CAF	Acerto
<b>6</b>	LID + CAF	LID + CAF	Acerto
<b>7</b>	LID + CAF	LID + CAF	Acerto
<b>8</b>	LID + CAF	LID + CAF	Acerto
<b>9</b>	LID + CAF	LID + CAF	Acerto
<b>10</b>	LID + CAF	LID + CAF	Acerto
<b>11</b>	LID + CAF	LID + CAF	Acerto
<b>12</b>	LID + CAF	LID + CAF	Acerto
<b>13</b>	COC + CAF	COC + CAF + LID	Possível erro
<b>14</b>	COC + CAF	COC + CAF	Acerto
<b>15</b>	COC + CAF	COC + CAF	Acerto
<b>16</b>	COC + CAF	COC + CAF + LID	Acerto
<b>17</b>	COC + CAF	COC + CAF + LID	Acerto
<b>18</b>	COC + CAF	COC + CAF + LID	Possível erro
<b>19</b>	COC + CAF	COC + CAF	Acerto
<b>20</b>	COC + CAF	COC + CAF + LID	Erro
<b>21</b>	COC + CAF	COC + CAF	Acerto
<b>22</b>	LID + COC	LID + COC	Acerto

<sup>7</sup> É importante ressaltar que, neste contexto, classificação é sinônimo de construção de um modelo que divida as amostras de acordo com padrões de comportamento desejados/esperados.

<b>Amostra</b>	<b>Mistura</b>	<b>Classificação</b>	<b>Resultado</b>
<b>23</b>	LID + COC	LID + COC	Acerto
<b>24</b>	LID + COC	LID + COC	Acerto
<b>25</b>	LID + COC	LID + COC	Acerto
<b>26</b>	LID + COC	COC + CAF + LID	Acerto
<b>27</b>	LID + COC	LID + COC	Acerto
<b>28</b>	LID + COC	LID + COC	Acerto
<b>29</b>	LID + COC	LID + COC	Acerto
<b>30</b>	LID + COC	LID + COC	Acerto
<b>31</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>32</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>33</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>34</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>35</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>36</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>37</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>38</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>39</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>40</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>41</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>42</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>43</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>44</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>45</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>46</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>47</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>48</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>49</b>	COC + CAF + LID	COC + CAF + LID	Acerto

<b>Amostra</b>	<b>Mistura</b>	<b>Classificação</b>	<b>Resultado</b>
<b>50</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>51</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>52</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>53</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>54</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>55</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>56</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>57</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>58</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>59</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>60</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>61</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>62</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>63</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>64</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>65</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>66</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>67</b>	COC + CAF + LID	COC + CAF + LID	Acerto
<b>68</b>	COC + CAF + LID	COC + CAF + LID	Acerto

O modelo desenvolvido permitiu uma correta classificação de, pelo menos, 95,6% das amostras. As amostras 13 e 18, de composição cocaína:caféina (8:2) e (1:9), respectivamente, são vistas como potenciais erros, pois não estão tocando na aresta, porém o erro associado à classificação pode ser menor que o erro do método, fazendo com que elas sejam classificadas corretamente. A sua composição pode revelar uma fragilidade no método, quando a



concentração de cocaína ou cafeína, em uma mistura binária, for muito baixa (em torno de 20% para a cocaína e 10% para a cafeína).

#### 6.5.4 *Análise de loadings*

A análise do gráfico de loadings em relação à primeira Componente Principal revelou uma característica importante das transformações por derivada: uniformizar a importância das variáveis no modelo construído (Figura 33), o que dificulta a interpretação do gráfico de *loadings*, pois todos os máximos das bandas dos espectros são igualados a zero. O mesmo pode ser visto no gráfico 2-D de *loadings* PC1 vs. PC2, com os dados pré-processados por primeira derivada (Figura 34).

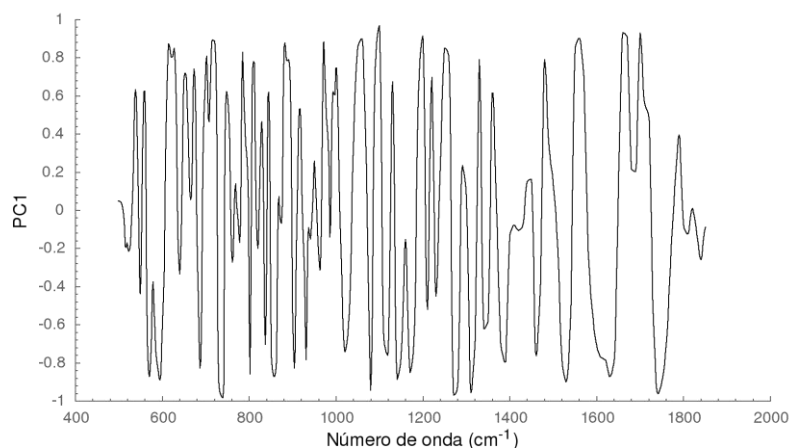


Figura 33 – Gráfico de *loadings* da PC1 dos dados pré-processados por auto escalamento e transformação por primeira derivada, refletindo a uniformidade de importância das variáveis no modelo de PCA construído.

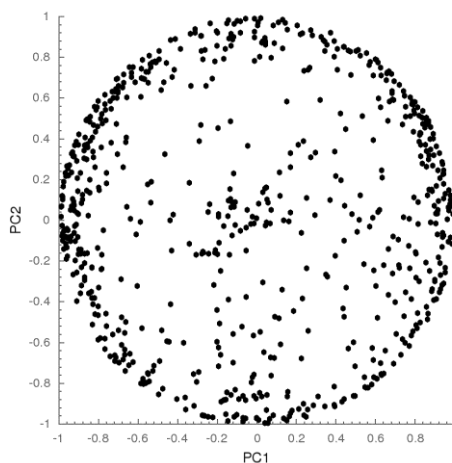


Figura 34 - Gráfico de *loadings* PC1 vs. PC2 para os dados pré-processados por primeira derivada

## 7 CONCLUSÃO

Através deste trabalho, pôde-se concluir que é possível classificar amostras de cocaína, lidocaína e cafeína, em misturas binárias e ternárias, de acordo com a estrutura dos dados (variações na composição). Não houve prejuízos ao modelo em se usar o espectro médio das amostras analisadas, embora eles não refletissem toda a dispersão contida nas replicatas, tampouco em utilizar apenas a região de  $1850\text{-}500\text{ cm}^{-1}$  para a construção do modelo.

Foi possível explicar 90,98% da variância residual total do sistema, com 3 componentes principais, para o modelo construído a partir dos dados pré-processados apenas por centralização pela média e escalamento por variância. Entretanto, quando submetidos à transformação por primeira derivada, o número ótimo de componentes principais diminuiu para 2, assim como a variância explicada total do sistema, indo para 72,82%. O pré-processamento por primeira derivada permitiu a construção de um modelo classificatório mais fiel à real composição das amostras, porém dificultou a interpretação dos *loadings* do sistema analisado.

## **8      PERSPECTIVAS FUTURAS**

As perspectivas futuras englobam: quantificar as amostras do sistema por PLS2, MCR-ALS e PCR, validar o método desenvolvido e reproduzir o mesmo modelo em equipamentos de infravermelho com diferentes técnicas de detecção.

## REFERÊNCIAS

- BERNARDO, N. P. *et al.* **Caffeine and other adulterants in seizures of street cocaine in Brazil.** International Journal of Drug Policy. 2003. Volume 14. p. 331-334.
- BOVENS, M. *et al.* **Chemometrics in forensic chemistry – Part I: Implications to the forensic workflow.** Forensic Science International. 2019. Volume 301. p. 82-90.
- CHALMERS, J. M.; EDWARDS, H. G. M.; HARGREAVES, M. D. **Infrared and Raman Spectroscopy in Forensic Science.** 1<sup>st</sup> Edition. United Kingdom. Wiley. 2012.
- COATES, J. **Interpretation of Infrared Spectra, A Practical Approach.** In: **Encyclopedia of Analytical Chemistry.** 1<sup>st</sup> Edition. Newtown. R. A. Meyers. 2006. p. 1-23.
- ELIAERTS, J. *et al.* **Rapid classification and quantification of cocaine in seized powders with ATR-FTIR and chemometrics.** Drug Testing and Analysis. 2017. Volume 9. p. 1480-1489.
- ESBENSEN, K. H.; SWARBRICK, B. **Multivariate Data Analysis: An introduction to Multivariate Analysis, Process Analytical Technology and Quality by Design.** 6<sup>th</sup> Edition. United States. CAMO Software Inc. 2018.
- GEMPERLINE, P. J. **Principal Component Analysis.** In: **Practical Guide to Chemometrics.** 2<sup>nd</sup> Edition. Florida. Taylor & Francis Group, LLC. 2006. p. 69-102.
- GROBÉRIO, T. S. *et al.* **Quantification of cocaine hydrochloride in seized drug samples by Infrared Spectroscopy and PLSR.** Journal of Brazilian Chemical Society. 2014. Volume 25. Número 9. p. 1696-1703.
- HESPANHOL, M. C.; PASQUINI, C.; MALDANER, A. O. **Evaluation of a low-cost portable near-infrared spectrophotometer for in situ cocaine profiling.** Talanta. 2019. Volume 200. p. 553-561.
- HOTELLING, H. **Analysis of a complex of statistical variables into Principal Components.** The Journal of Educational Psychology. 1933. Volume 24. Número 6. p. 417-441.
- JOLLIFFE, I. T. **Principal Component Analysis.** 2<sup>nd</sup> Edition. United Kingdom. Springer. 2002.
- KATON, J. E.; MCKITTRICK, P. T. **Infrared and Raman Group Frequencies of Cyclic Imides.** Applied Spectroscopy. 1990. Volume 44. Número 5. p. 812-817.
- LAPACHINSKE, S. F. *et al.* **Analysis of cocaine and its adulterants in drug for international trafficking seized by the Brazilian Federal Police.** Forensic Science International. 2015. Volume 247. p. 48-53.
- LEVINE, I. N. **Spectroscopy and Photochemistry.** In.: **Physical Chemistry.** 6<sup>th</sup> Edition. New York. McGraw-Hill. 2009.

- MAHARAJ, R. **Quantitative Analysis of Cocaine using Fourier Transform Infrared Spectroscopy-Attenuated Total Reflectance: A preliminary investigation**. The Internet Journal of Third World Medicine. 2008. Volume 7. Número 2. p. 1-5.
- MALDANER, A. O. *et al.* **Chemical profiling of street Cocaine from different Brazilian regions**. Journal of Brazilian Chemical Society. 2016. Volume 27. Número 4. p.718-726.
- MARCELO, M. C. A.; *et al.* **Profiling cocaine by ATR-FTIR**. Forensic Science International. 2015. Volume 246. p. 65-71.
- MATERAZZI, S. *et al.* **Cocaine profiling: Implementation of a predictive model by ATR-FTIR coupled with chemometrics in forensic chemistry**. Talanta. 2017. Volume 166. p. 328-335.
- MONTALVÃO, V. T. K. **Metodologia de calibração multivariada, associada á técnica de ATR-FTIR em tempo real, para o monitoramento quantitativo da síntese de biodiesel com diferentes catalisadores**. 2013. Trabalho de conclusão de curso (Graduação em Química com Atribuições Tecnológicas). Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro – RJ, 2013.
- MURO, C. K.; DOTY, K. C.; BUENO, J.; HALÁMKOVÁ, L.; LEDNEV, I. K. **Forensic Applications of Vibrational Spectroscopy**. In: **Forensic Science: A Multidisciplinary Approach**. 1<sup>st</sup> Edition. Weinheim, Germany. Wiley. 2016.
- OLIVEIRA, A. L. B. de. **Análise Quantitativa por Ressonância Magnética Nuclear e Quimiometria de misturas de amidas e aminas aromáticas**. 2011. Dissertação (Mestrado em Ciências em Química). Instituto Militar de Engenharia, Rio de Janeiro – RJ, 2011.
- PEARSON, K. **On lines and planes of closest fit to systems of points in space**. Philosophical Magazine. 1901. Volume 2. p. 559-572.
- PENIDO, C. A. F. de O. *et al.* **Quantification of cocaine in ternary mixtures using Partial Least Squares Regression applied to Raman and Fourier Transform Infrared Spectroscopy**. Journal of Raman Spectroscopy. 2017. p. 1-12.
- PERKINELMER. **FT-IR Spectroscopy: Attenuated Total Reflectance (Technical Note)**. USA. PerkinElmer Life and Analytical Sciences. 2005.
- RAMASAMY, R. **Vibrational Spectroscopic Studies of Imidazole**. Armenian Journal of Physics. 2015. Volume 8. Número 1. p. 51-55.
- RODRIGUES, N. V. S. *et al.* **Analysis of seized Cocaine samples by using Chemometric Methods and FTIR Spectroscopy**. Journal of Brazilian Chemical Society. Brasil. 2013. Volume 24. Número 3. p. 507-517.
- ROGGO, Y. *et al.* **A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies**. Journal of Pharmaceutical and Biomedical Analysis. 2007. Volume 44. p. 683-700.

SANT'ANA, L. D. *et al.* **Evaluation of cocaine samples seized in the streets of the state of Rio de Janeiro, Brazil.** Química Nova. 2019. Volume 42. Número 4. p. 379-386.

SILVA, A. F. da. **Discriminação e Quantificação de cocaína e adulterantes em drogas por Espectroscopia de Infravermelho e Resolução Multivariada de curvas.** 2016. Dissertação (Mestrado em Tecnologia Química e Biológica). Instituto de Química, Universidade de Brasília, Brasília- DF, 2016.

SILVERSTEIN, R. M.; WEBSTER, F. X.; KIEMLE, D. J. **Spectrometric Identification of Organic Compounds.** 7<sup>th</sup> Edition. New York. John Wiley & Sons, Inc. 2005. p. 72-126.

SKOOG, D. A.; WEST, D. M.; HOLLER, F. J.; CROUCH, S. R. **Fundamentals of Analytical Chemistry.** 9<sup>th</sup> Edition. Belmont, California. Brooks/Cole, Cengage Learning. 2014.

SOUZA, L. M. de *et al.* **A survey of adulterants used to cut cocaine in samples seized in the Espírito Santo State by GC-MS allied to chemometric tools.** Science and Justice. 2016. Volume 56. p. 73-79.

SWARBRICK, B. **Multivariate Data Analysis for Dummies.** CAMO Software Special Edition. England. Wiley. 2012.

SWGDRUG. **Scientific Working Group for the Analysis of Seized Drugs.** Version 7.0. United States. SWGDRUG. Ago. 2014.

UNODC. **Global overview of drug demand and supply.** Vienna – Austria. United Nations. June 2019. World Drug Report 2019. Booklet 2.

UNODC. **Drugs and age: Drugs and associated issues among young people and older people.** Vienna – Áustria. United Nations. June 2018. World Drug Report 2018. Booklet 4.

UNODC. **Recommended methods for the Identification and Analysis of Cocaine in Seized Materials:** Manual for use by National Drug Analysis Laboratories. 1<sup>st</sup> Edition. New York - United States. 2012.

WOLD, S.; ESBENSEN, K.; GELADI, P. **Principal Component Analysis.** Chemometrics and Intelligent Laboratory Systems. 1987. Volume 2. p. 37-52.

YOUNG, H. D.; FREEDMAN, R. A.; FORD, A. F. **Electromagnetic Waves.** In: **University Physics with Modern Physics.** 13<sup>th</sup> Edition. San Francisco, California. Pearson. 2012.